PA 178512

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

November 24, 1999

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: *60/095,653*
FILING DATE: *August 07, 1998*

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

N. WOODSON
Certifying Officer

A/Pros

PROVISIONAL APPLICATION FOR PATENT
COVER SHEET

Case No. GENSET.037PR
Date: August 7, 1998
Page 1

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

ATTENTION: PROVISIONAL PATENT APPLICATION

Sir:

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR § 1.53(b)(2).

For: **NUCLEIC ACIDS ENCODING HUMAN TBC-1 PROTEIN AND POLYMORPHIC MARKERS THEREOF**

Name of First Inventor: Lydie Bougueleret
Residence Address: 108, avenue Victor Hugo, 92170 Vanves, France

Name of Second Inventor: Marta Blumenfeld
Residence Address: 5, rue Tagore, 75013 Paris, France

Name of Third Inventor: Ilya Chumakov
Residence Address: 16, rue des Chevrefeuilles, 7700 Vaux-le-Penil, France

Enclosed are:

(X) Specification in 128 pages.

(X) Sequence Listing in 37 pages.

(X) Sequence Submission in 1 page.

(X) Sequence Listing in computer readable form.

(X) 4 sheet(s) of drawings.

(X) A check in the amount of $150 to cover the filing fee is enclosed.

(X) A return prepaid postcard.

(X) The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Account No. 11-1410. A duplicate copy of this sheet is enclosed.

Was this invention made by an agency of the United States Government or under a contract with an agency of the United States Government?

(X) No.
( ) Yes. The name of the U.S. Government agency and the Government contract number are:

KNOBBE, MARTENS, OLSON & BEAR, LLP
620 NEWPORT CENTER DR 16TH FLOOR NEWPORT BEACH, CA 92660
(949) 760-0404 FAX (949) 760-9502

(X) Please send correspondence to:

Daniel Hart
Knobbe, Martens, Olson & Bear, LLP
620 Newport Center Dr., 16th Floor
Newport Beach, CA 92660

Respectfully submitted,

Daniel Hart
Registration No. 40,637

S:\DOCS\DOH\DOH-1933.DOC
080798

INTELLECTUAL PROPERTY LAW

KNOBBE, MARTENS, OLSON & BEAR

A LIMITED LIABILITY PARTNERSHIP INCLUDING PROFESSIONAL CORPORATIONS

PATENT, TRADEMARK AND COPYRIGHT CAUSES

501 WEST BROADWAY
SUITE 1400
SAN DIEGO, CALIFORNIA 92101-3505
(619) 235-8550
FAX (619) 235-0176
INTERNET: WWW.KMOB.COM

08/07/98

Assistant Commissioner for Patents
Washington, D.C. 20231

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

Attorney Docket No.	:	GENSET.037PR
Applicant(s)	:	Bougueleret, et al.
For	:	NUCLEIC ACIDS ENCODING HUMAN TBC-1 PROTEIN AND POLYMORPHIC MARKERS THEREOF
Attorney	:	Daniel Hart
"Express Mail" Mailing Label No.	:	EL049129612US
Date of Deposit	:	August 7, 1998

I hereby certify that the accompanying

Transmittal in Duplicate; Specification in 128 pages; 4 sheets of drawings; Sequence Listing in 37 pages; Sequence Submission in 1 page; Sequence Listing in computer readable form; $150 Check for Filing Fee(s); Return Prepaid Postcard

are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and are addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.

John J. Riedel

S:\DOCS\DOH\DOH-1935.DOC
080798

275 BATTERY STREET
SUITE 1840
SAN FRANCISCO, CALIFORNIA 94111
(415) 954-4114
FAX (415) 954-4111

620 NEWPORT CENTER DRIVE
SIXTEENTH FLOOR
NEWPORT BEACH, CALIFORNIA 92660
(949) 760-0404
FAX (949) 760-9502

3801 UNIVERSITY AVENUE
SUITE 710
RIVERSIDE, CALIFORNIA 92501
(909) 781-9231
FAX (909) 781-4507

* A PROFESSIONAL CORPORATION
† ALSO MEMBER OF D.C. BAR
†† ALSO BARRISTER AT LAW (U.K.)
** U.S. PATENT AGENT

TITLE OF THE INVENTION : Nucleic acids encoding human TBC-1 protein and polymorphic markers thereof.

FIELD OF THE INVENTION

The invention concerns genomic and cDNA sequences of the human *TBC-1* gene, a biallelic marker of the *TBC-1* gene and the association established between this marker and prostate cancer. The invention provides means to determine the predisposition of individuals to prostate cancer as well as means for the diagnosis of this cancer and for the prognosis/detection of an eventual treatment response to therapeutic agents acting against prostate cancer.

BACKGROUND OF THE INVENTION

Prostate cancer

The incidence of prostate cancer has dramatically increased over the last decades. It averages 30-50/100,000 males in Western European countries as well as within the US White male population. In these countries, it has recently become the most commonly diagnosed malignancy, being one of every four cancers diagnosed in American males. Prostate cancer's incidence is very much population specific, since it varies from 2/100,000 in China, to over 80/100,000 among African-American males.

In France, the incidence of prostate cancer is 35/100,000 males and it is increasing by 10/100,000 per decade. Mortality due to prostate cancer is also growing accordingly. It is the second cause of cancer death among French males, and the first one among French males aged over 70. This makes prostate cancer a serious burden in terms of public health.

Prostate cancer is a latent disease. Many men carry prostate cancer cells without overt signs of disease. Autopsies of individuals dying of other causes show prostate cancer cells in 30 % of men at age 50 and in 60 % of men at age 80. Furthermore, prostate cancer can take up to 10 years to kill a patient after the initial diagnosis.

The progression of the disease usually goes from a well-defined mass within the prostate to a breakdown and invasion of the lateral margins of the prostate, followed by metastasis to regional lymph nodes, and metastasis to the bone marrow. Cancer metastasis to bone is common and often associated with uncontrollable pain.

Unfortunately, in 80 % of cases, diagnosis of prostate cancer is established when the disease has already metastasized to the bones. Of special interest is the observation that prostate cancers frequently grow more rapidly in sites of metastasis than within the prostate itself.

Early-stage diagnosis of prostate cancer mainly relies today on Prostate Specific Antigen (PSA) dosage, and allows the detection of prostate cancer seven years before clinical symptoms become apparent. The effectiveness of PSA dosage diagnosis is however limited, due to its inability to discriminate between malignant and non-malignant affections of the organ and because not all prostate cancers give rise to an elevated serum PSA concentration. Furthermore, PSA dosage and other currently available approaches such as physical examination, tissue biopsy and bone scans are of limited value in predicting disease progression.

Therefore, there is a strong need for a reliable diagnostic procedure which would enable a more systematic early-stage prostate cancer prognosis.

Although an early-stage prostate cancer prognosis is important, the possibility of measuring the period of time during which treatment can be deferred is also interesting as currently available medicaments are expensive and generate important adverse effects. However, the aggressiveness of prostate tumors varies widely. Some tumors are relatively aggressive, doubling every six months whereas others are slow-growing, doubling once every five years. In fact, the majority of prostate cancers grows relatively slowly and never becomes clinically manifest. Very often, affected patients are among the elderly and die from another disease before prostate cancer actually develops. Thus, a significant question in treating prostate carcinoma is how to discriminate between tumors that will progress and those that will not progress during the expected lifetime of the patient.

Hence, there is also a strong need for detection means which may be used to evaluate the aggressiveness or the development potential of prostate cancer tumors once diagnosed.

Furthermore, at the present time, there is no means to predict prostate cancer susceptibility. It would also be very beneficial to detect individual susceptibility to prostate cancer. This could allow preventive treatment and a careful follow up of the development of the tumor.

A further consequence of the slow growth rate of prostate cancer is that few cancer cells are actively dividing at any one time, rendering prostate cancer generally resistant to radiation and chemotherapy. Surgery is the mainstay of treatment but it is largely ineffective and removes the ejaculatory ducts, resulting in impotence. Oral oestrogens and luteinizing releasing hormone analogs are also used for treatment of prostate cancer. These hormonal treatments provide marked improvement for many patients, but they only provide temporary relief. Indeed, most of these cancers soon relapse with the development of hormone-resistant tumor cells and the oestrogen treatment can lead to serious cardiovascular complications. Consequently, there is a strong need for preventive and curative treatment of prostate cancer.

Efficacy/tolerance prognosis could be precious in prostate cancer therapy. Indeed, hormonal therapy, the main treatment currently available, presents important side effects. The use of chemotherapy is limited because of the small number of patients with chemosensitive tumors. Furthermore the age profile of the prostate cancer patient and intolerance to chemotherapy make the systematic use of this treatment very difficult.

Therefore, a valuable assessment of the eventual efficacy of a medicament to be administered to a prostate cancer patent as well as the patent's eventual tolerance to it may permit to enhance the benefit/risk ratio of prostate cancer treatment.

It is known today that there is a familial risk of prostate cancer. Clinical studies in the 1950s had already demonstrated a familial aggregation in prostate cancer. Control-case clinical studies have been conducted more recently to attempt to evaluate the incidence of the genetic risk factors in the disease. Thus Steinberg et al., 1990, and McWhorter et al., 1992 confirm that the risk of prostate cancer is increased in subjects having one or more relatives already affected by the disease and when forms of early diagnosis in the relatives exist.

It is now well established that cancer is a disease caused by the deregulation of the expression of certain genes. In fact, the development of a tumor necessitates an important succession of steps. Each of these steps comprises the deregulation of an important gene intervening in the normal metabolism of the cell and the emergence of an abnormal cellular sub-clone which overwhelms the other cell types because of a proliferative advantage. The genetic origin of this concept has found confirmation in the isolation and the characterization of genes which could be responsible. These genes, commonly called "cancer genes", have an important role in the

normal metabolism of the cell and are capable of intervening in carcinogenesis following a change.

Recent studies have identified three groups of genes which are frequently mutated in cancer. The first group of genes, called oncogenes, are genes whose products activate cell proliferation. The normal non-mutant versions are called protooncogenes. The mutated forms are excessively or inappropriately active in promoting cell proliferation, and act in the cell in a dominant way in that a single mutant allele is enough to affect the cell phenotype. Activated oncogenes are rarely transmitted as germline mutations since they may probably be lethal when expressed in all the cells. Therefore oncogenes can only be investigated in tumor tissues.

The second group of genes which are frequently mutated in cancer, called tumor suppressor genes, are genes whose products inhibit cell growth. Mutant versions in cancer cells have lost their normal function, and act in the cell in a recessive way in that both copies of the gene must be inactivated in order to change the cell phenotype. Most importantly, the tumor phenotype can be rescued by the wild type allele, as shown by cell fusion experiments first described by Harris and colleagues (1969). Germline mutations of tumor suppressor genes may be transmitted and thus studied in both constitutional and tumor DNA from familial or sporadic cases. The current family of tumor suppressors includes DNA-binding transcription factors (i.e., p53, WT1), transcription regulators (i.e., RB, APC, probably BRCA1), protein kinase inhibitors (i.e., p16), among others (for review, see Haber D & Harlow E, 1997).

The third group of genes which are frequently mutated in cancer, called mutator genes, are responsible for maintaining genome integrity and/or low mutation rates. Loss of function of both alleles increases cell mutation rates, and as a consequence, proto-oncogenes and tumor suppressor genes may be mutated. Mutator genes can also be classified as tumor suppressor genes, except for the fact that tumorigenesis caused by this class of genes cannot be suppressed simply by restoration of a wild-type allele, as described above. Genes whose inactivation may lead to a mutator phenotype include mismatch repair genes (i.e., MLH1, MSH2), DNA helicases (i.e., BLM, WRN) or other genes involved in DNA repair and genomic stability (i.e., p53, possibly BRCA1 and BRCA2) (For review see Haber D & Harlow E, 1997; Fishel R & Wilson T. 1997; Ellis NA,1997).

There is growing evidence that a critical event in the progression of a tumor cell from a non-metastatic to metastatic phenotype is the loss of function of metastasis-suppressor genes.

These genes specifically suppress the ability of a cell to metastasize. Work from several groups has demonstrated that human chromosomes 8, 10, 11 and 17 encode prostate cancer metastasis suppressor activities. However, other human chromosomes such as chromosomes 1, 7, 13, 16, and 18 may also be associated to prostate cancer.

It thus remains to localize and to identify the genes specifically involved in the development and the progression of prostate cancers starting from the genetic analysis of the hereditary and the non-hereditary forms and to define their clinical implications in terms of prognosis and therapeutic innovations.

SUMMARY OF THE INVENTION

The present invention concerns a human *TBC-1* genomic sequence which is included in a previously unknown candidate region of prostate cancer located on chromosome 4. The invention also concerns a biallelic marker belonging to the *TBC-1* gene which is strongly associated to prostate cancer.

Human *TBC-1* genomic sequences, corresponding cDNA and *TBC-1* coding or regulatory sequences

The human gene encoding a human TBC-1 polypeptide has been found by the inventors to be located on human chromosome 4. The inventors have isolated and sequenced a portion of the genomic sequence of the human *TBC-1* gene. Thus, a first object of the present invention relates to a *TBC-1* genomic sequence comprising the first three exons and the 5' regulatory region.

The expressions "*TBC-1* gene" or "TBC-1 protein" are herein intended to designate the human *TBC-1* gene or the human TBC-1 protein.

The present invention concerns a purified or isolated nucleic acid encoding a human TBC-1 protein, wherein said TBC-1 protein comprises an amino acid sequence of SEQ ID No 5, a nucleotide sequence complementary thereto, a fragment or a variant thereof.

As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude

is expressly contemplated. As an example, purification from 0.1 % concentration to 10 % concentration is two orders of magnitude.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment.

The present invention concerns a purified or isolated nucleic acid comprising a nucleotide sequence of SEQ ID No 1, a nucleotide sequence complementary thereto, a fragment or a variant thereof.

The invention also concerns a purified or isolated nucleic acid comprising at least 20 consecutive nucleotides of the nucleotide sequence SEQ ID No 1, or a nucleotide sequence complementary thereto.

The invention further deals with a purified or isolated nucleic acid comprising the nucleotide sequence of a regulatory region which is located upstream of the first exon of the *TBC-1* gene and which is contained in the *TBC-1* genomic sequence of SEQ ID No 1. The invention thus encompasses a purified or isolated nucleic acid comprising the nucleotide sequence of SEQ ID No 2, or a sequence complementary thereto or biologically active fragment or variant thereof, as well as any sequence of 8 to about 2000 consecutive nucleotides, preferably of 10 to 500 consecutive nucleotides, more preferably 10 to 300, included therein. More particularly, the invention further includes specific elements within this regulatory region, these elements preferably including the promoter region.

The present invention is also directed to a polynucleotide comprising a functional portion of a regulatory region contained in the contemplated *TBC-1* genomic sequence and to its use in a recombinant expression vector carrying a polynucleotide encoding a polypeptide or a nucleic acid of interest.

The transcription of the genomic sequence leads to more than one mRNA final product, due to alternative splicing events. Two mRNAs have been identified and they result from two distinct first exons, namely Exon 1 and Exon 1*bis*. Therefore, the present invention concerns a purified or isolated nucleic acid comprising a nucleotide sequence selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4, a nucleotide sequence complementary thereto or a variant thereof.

The invention also concerns a purified or isolated nucleic acid having at least 95% nucleotide identity with a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4, or a nucleotide sequence complementary thereto.

The invention also deals with a purified or isolated nucleic acid comprising at least 20 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4.

A further object of the invention consists in polynucleotide fragments of the *TBC-1* genomic sequence that are useful for detecting the presence of an unaltered or altered copy of this gene within the human genome of a given individual and also for the detection and/or quantification of the expression of the *TBC-1* gene in said individual host organism.

When used herein, an altered copy of the *TBC-1* gene according to the invention is intended to designate the *TBC-1* gene that has undergone at least one substitution or deletion of one or several nucleotides, wherein said nucleotide substitution, addition or deletion of one or several nucleotides causes a change in the amino acid sequence of SEQ ID No 5 or alternatively causes an increase or a decrease in the expression of the *TBC-1* gene.

When used herein, the term "*TBC-1* gene" is intended to define an entity which can comprise some or all the following elements : exons, introns, promoter, regulatory regions, 5'UTR, 3' UTR and regions never transcribed and located either upstream or downstream of the coding sequence of *TBC-1*.

Another aspect of the present invention is a purified and/or isolated *TBC-1* genomic sequence comprising at least one of the biallelic polymorphisms according to the present invention, a sequence complementary thereto, a fragment or a variant thereof. In a preferred embodiment,

the *TBC-1* gene comprises one of the nucleotide sequences of SEQ ID Nos 7 and 8, a sequence complementary thereto, a fragment or a variant thereof.

Biallelic markers

The invention also concerns a purified and/or isolated biallelic marker located in the sequence of the *TBC-1* gene, preferably a biallelic marker comprising an allele associated with prostate cancer, with an early onset of prostate cancer, with a response to a prophylactic or therapeutic agent administered for cancer treatment, particularly prostate cancer, with the level of aggressiveness of prostate cancer tumors, with a modified or forthcoming expression of the *TBC-1* gene, with a modified or forthcoming production of the TBC-1 protein, or with the production of a modified TBC-1 protein.

As used herein, the term "aggressiveness" of prostate cancer tumors refers to the metastatic potential of these tumors.

The invention also relates to a nucleotide sequence, preferably a purified and/or isolated polynucleotide comprising a sequence defining a biallelic marker located in the sequence of the *TBC-1* gene, particularly the genomic sequence of SEQ ID No 1, a fragment or variant thereof or a sequence complementary thereto. As used herein, the terminology "defining a biallelic marker" means that a sequence includes a polymorphic base from a biallelic marker. The sequences defining a biallelic marker may be of any length consistent with their intended use, provided that they contain a polymorphic base from a biallelic marker. The sequence has between 1 and 500 nucleotides in length, preferably between 5, 10, 15, 20, 25 or 40 and 200 nucleotides and more preferably between 30 and 50 nucleotides in length. Preferably, the sequences defining a biallelic marker include the polymorphic base of one of SEQ ID Nos 7 and 8. In some embodiments the sequences defining a biallelic marker comprise one of the sequences selected from the group consisting of SEQ ID Nos 7 and 8. Likewise, the term "marker" or "biallelic marker" requires that the sequence is of sufficient length to practically (although not necessarily unambiguously) identify the polymorphic allele, which usually implies a length of at least 4, 5, 6, 10, 15, 20, 25 or 40 nucleotides.

The invention further concerns a nucleic acid encoding a TBC-1 protein, wherein said nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID Nos 7 and 8.

The invention also relates to nucleotide sequence selected from the group consisting of SEQ ID Nos 7 and 8 or a fragment or a variant thereof.

The invention also pertains to a nucleotide sequence selected from the group consisting of a variant or fragment of SEQ ID Nos 7 and 8, said fragment comprising at least 8 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID Nos 7 and 8 and including the polymorphic base thereof.

Identification and characterization of further biallelic markers

Another aspect of the present invention is a method for identifying biallelic markers in the genomic region harboring the *TBC-1* gene. The method comprises the steps of :

- obtaining a nucleic acid containing the genomic region harboring the *TBC-1* gene, preferably a nucleic acid comprising the nucleotide sequence of SEQ ID No 1 or a nucleic acid comprising a portion of a nucleotide sequence selected from the group consisting of SEQ ID Nos 3 and 4 that is present in the *TBC-1* genomic sequence of SEQ ID No 1, or fragments or variants thereof;
- sequencing at least a portion of said nucleic acid containing the genomic region harboring the *TBC-1* genomic sequence to determine a plurality of primer sequences capable of amplifying portions of said genomic region harboring the *TBC-1* gene;
- amplifying portions of the genomic region harboring the *TBC-1* genomic sequence from a plurality of individuals using said primers to obtain a plurality of amplicons; and
- sequencing said plurality of amplicons to identify biallelic markers in the genomic region harboring the *TBC-1* genomic sequence.

The invention also concerns a method for the identification and characterization of a biallelic marker in linkage disequilibrium with a biallelic marker of the *TBC-1* gene, preferably a biallelic marker of the *TBC-1* gene of which one allele is associated with a trait, preferably with prostate cancer. In one embodiment, the biallelic marker of the *TBC-1* gene is in the genomic region harboring the *TBC-1* gene, but outside of the *TBC-1* gene itself. In another embodiment, the biallelic marker in linkage disequilibrium with a biallelic marker of the *TBC-1* gene is itself located within the *TBC-1* gene.

The method comprises the following steps:

- providing a first biallelic marker of the *TBC-1* gene;
- amplifying a position of the genomic region harboring the *TBC-1* gene;
- sequencing and identifying second biallelic markers in said amplified portion;

- conducting a linkage disequilibrium analysis between said first biallelic marker and second biallelic markers; and
- identifying second biallelic markers in linkage disequilibrium with said first marker.

In a preferred embodiment, the step of sequencing and identifying second biallelic markers comprises sequencing second biallelic markers within the *TBC-1* gene.

Once identified, the sequences in linkage disequilibrium with a biallelic marker of the *TBC-1* gene may be used in any of the methods described herein, including methods for determining an association between a biallelic marker and a trait, methods for identifying individuals having a predisposition for a trait, methods of administration of prophylactic or therapeutic agents disease treatment, methods of identifying individuals likely to respond positively or negatively to said agents, and methods of using drugs and vaccines.

Oligonucleotide probes and primers

The invention relates also to oligonucleotide molecules useful as probes or primers, wherein said oligonucleotide molecules hybridize specifically with a nucleotide sequence comprised in the *TBC-1* gene, particularly the *TBC-1* genomic sequence of SEQ ID No 1. More particularly, the present invention also concerns oligonucleotides for the detection of alleles of biallelic markers of the *TBC-1* gene, preferably those associated with prostate cancer, with an early onset of prostate cancer, with a susceptibility to prostate cancer, with the level of aggressiveness of prostate cancer tumors, with a modified or forthcoming expression of the *TBC-1* gene, with a modified production of the TBC-1 protein, or with the production of a modified TBC-1 protein. These oligonucleotides are characterized in that they can hybridize with a *TBC-1* gene, preferably with a polymorphic *TBC-1* gene and more preferably with a region of a *TBC-1* gene comprising a polymorphic site of a specific allele associated with prostate cancer, with the level of aggressiveness of prostate cancer tumors or with modifications in the regulation of the expression of the *TBC-1* gene. These oligonucleotides are useful either as primers for use in various processes such as DNA amplification and microsequencing or as probes for DNA recognition in hybridization analyses.

In a first embodiment, a purified or isolated nucleic acid useful as an amplification primer or as a probe of the invention comprises a polynucleotide selected from the group consisting of the nucleotide sequences complementary to any sequence of a strand of the *TBC-1* gene and the nucleotide sequences complementary to any one of the SEQ ID Nos 2-4.

A nucleic acid probe or primer according to the invention comprises at least 8 consecutive nucleotides of a polynucleotide of SEQ ID No 1, preferably from 8 to 200 consecutive nucleotides, more particularly from 10, 15, 20 or 30 to 100 consecutive nucleotides, more preferably from 10 to 50 nucleotides, and most preferably from 40 to 50 consecutive nucleotides of a polynucleotide of SEQ ID No 1.

Preferred probes or primers of the invention comprise at least 20 consecutive nucleotides of a polynucleotide selected from the group consisting of : (a) the nucleic acid beginning at the nucleotide in position 1999 and ending at the nucleotide in position 13248 of SEQ ID No 1; (b) the nucleic acid of SEQ ID No 3; (c) the nucleic acid of SEQ ID No 4.

Preferred probes or primers of the invention comprise at least 8 consecutive nucleotides of a polynucleotide consisting of the nucleotide sequence beginning at the nucleotide in position 1 and ending at the nucleotide in position 2000 of the nucleotide sequence of SEQ ID No 1.

In some embodiments, the oligonucleotides comprise the polymorphic base of a sequence selected from SEQ ID Nos 7-8. In other embodiments, the oligonucleotides, preferably those selected from SEQ ID Nos 11-12, have a 3' terminus immediately adjacent to a polymorphic base in the *TBC-1* gene, such as a polymorphic base in one of SEQ ID Nos 7-8. In other embodiments, the oligonucleotide is capable of discriminating between different alleles of a biallelic marker in the *TBC-1* gene, including the biallelic markers of SEQ ID Nos 7-8. For example, the oligonucleotide may be capable of specifically hybridizing to one allele of a biallelic marker, including one of the biallelic markers of SEQ ID Nos 7-8.

In a first preferred embodiment, the probe or primer is suspended in a suitable buffer in view of performing a hybridization or an amplification reaction.

In a second embodiment, the oligonucleotide probe, which may be immobilized on a support, is capable of hybridizing with a *TBC-1* gene, particularly with the genomic sequence of SEQ ID No 1, preferably with a region of the *TBC-1* gene which comprises a biallelic marker of the present invention. The techniques for immobilizing a nucleotide primer or probe on a solid support are well-known to the skilled artisan and include, but are not limited to, the immobilization techniques described in the present application.

In a third embodiment, the primer is complementary to any nucleotide sequence of the *TBC-1* gene, particularly of the *TBC-1* genomic sequence of SEQ ID No 1, and can be used to amplify a region of the *TBC-1* gene contained in the nucleic acid sample to be tested which includes a polymorphic base of at least one biallelic marker according to the present invention. Preferably, the amplified region includes a polymorphic base of the biallelic marker defining by SEQ ID Nos 7-8.

In a fourth embodiment, the oligonucleotides of the invention can hybridize with at least a portion of an intron or of the regulatory sequences of the *TBC-1* gene. Particularly preferred oligonucleotides of the invention hybridize with a sequence comprised in an intron or in the regulatory sequences of the *TBC-1* gene. In an other preferred embodiment, the oligonucleotides of the invention can hybridize with at least a portion of an exon selected in the group of exons 1, 1*bis* and 2. The length of oligonucleotides hybridizing, either partially or completely, with such regions of the *TBC-1* gene is between 1 and 100, preferably between 1 and 50, more preferably between 5 and 30 and even more preferably between 5 and 15 nucleotides.

The terms "sample" or "material sample" are used herein to designate a solid or a liquid material suspected to contain a polynucleotide or a polypeptide of the invention. A solid material may be, for example, a tissue slice or biopsy within which is searched the presence of a polynucleotide encoding a TBC-1 protein, either a DNA or RNA molecule or within which is searched the presence of a native or a mutated TBC-1 protein, or alternatively the presence of a desired protein of interest the expression of which has been placed under the control of a *TBC-1* regulatory polynucleotide. A liquid material may be, for example, any body fluid such as serum, urine etc., or a liquid solution resulting from the extraction of nucleic acid or protein material of interest from a cell suspension or from cells in a tissue slice or biopsy. The term "biological sample" is also used and is more precisely defined within the Section dealing with DNA extraction.

The nucleic acid probes and primers of the invention are also used to detect and/or amplify a portion of the *TBC-1* gene within which a mutation causes a change either in the expression level of the *TBC-1* gene or a change in the amino acid sequence of the *TBC-1* gene translation product.

Amplification of the *TBC-1* gene

The invention also concerns a method for the amplification of a region of the *TBC-1* gene, particularly the *TBC-1* genomic sequence of SEQ ID No 1, or a fragment or a variant thereof in a test sample. The method comprises the steps of :

- contacting a test sample suspected of containing the desired *TBC-1* sequence or portion thereof with amplification reaction reagents comprising a pair of amplification primers such as those described above, the primers being located on either side of the *TBC-1* nucleotide region to be amplified. The method may further comprise the step of detecting the amplification product. For example, the amplification product may be detected using a detection probe that can hybridize with an internal region of the amplicon sequences. Alternatively, the amplification product may be detected with any of the primers used for the amplification reaction themselves, optionally under a labeled form.

Identification of an association between a set of biallelic markers and a trait

It is another object of the present invention to provide a method for the identification and characterization of an association between an allele of one or more biallelic markers of a *TBC-1* gene and a trait. The method comprises the steps of :

- genotyping a marker or a group of biallelic markers according to the invention in trait positive and trait negative individuals; and
- establishing a statistically significant association between one allele of at least one marker and the trait.

Preferably, the trait positive and trait negative individuals are selected from non-overlapping phenotypes as regards to the trait under study. In one embodiment, the biallelic marker is defined by the sequences of SEQ ID Nos 7-8.

In a preferred embodiment, the trait is prostate cancer, an early onset of prostate cancer, a susceptibility to prostate cancer, the level of aggressiveness of prostate cancer tumors, a modified expression of the *TBC-1* gene, a modified production of the TBC-1 protein, or the production of a modified TBC-1 protein.

If the trait is a beneficial response or inversely a side effect to treatment of prostate cancer, the method of the invention referred to above further comprises some or all of the following steps :

- selecting a population or cohort of subjects diagnosed as suffering from prostate cancer;

- administering a specified treatment of prostate cancer to said cohort of subjects;
- monitoring the outcome of drug administration and identifying those individuals that are trait positive or trait negative relative to the treatment;
- taking from said cohort biological samples containing DNA and testing this DNA for the presence of a specific allele or of a set of alleles of biallelic markers of the *TBC-1* gene;
- analyzing the distribution of alleles of biallelic markers between trait positive and trait negative individuals; and
- performing a statistical analysis to determine a statistically significant association between the presence or absence of the specific allele or of a specific set of alleles of biallelic markers of the *TBC-1* gene and the treatment related trait.

The step of testing for and detecting the presence of DNA comprising specific alleles of a biallelic marker or a group of biallelic markers of the present invention can be carried out as described further below.

Identification of a trait causing mutation in the TBC-1 gene

A further embodiment of the invention is a method to identify a trait causing mutation in the *TBC-1* gene, particularly the genomic sequence of SEQ ID No 1, pursuant to the detection of an association between alleles of one or several of the biallelic markers of the present invention and a particular trait. The particular trait can in particular be prostate cancer, the level of aggressiveness of prostate cancer tumors, a modified expression of the *TBC-1* gene, a modified production of the TBC-1 protein, or the production of a modified TBC-1 protein.

A trait causing mutation is a mutation which is at least partly responsible for a particular detectable phenotype in an individual. The mutation may comprise point mutations, deletions, or insertions in the *TBC-1* gene, particularly in the genomic sequence of SEQ ID No 1. The mutations may lie within the coding sequence for TBC-1 protein or within intronic and/or regulatory regions in *TBC-1* gene, particularly in the genomic sequence of SEQ ID No 1, including splice sites, 5' UTRs, 3' UTRs and promoter sequences, including one or more transcription factor binding sites.

This method comprises the following steps :
- amplifying a region of the *TBC-1* gene, particularly of the *TBC-1* genomic sequence of SEQ ID No 1, from DNA samples of trait positive and trait negative individuals;
- sequencing the amplified region;

- comparing DNA sequences from trait positive and trait negative individuals; and
- determining mutations specific to trait positive patients.

In some embodiments, the amplified region is a region located close to a biallelic marker of *TBC-1* gene. In a further embodiment, the amplified region is located close to the biallelic marker defined by the sequences SEQ ID 7-8.

The invention also concerns a mutated *TBC-1* gene, particularly a mutated *TBC-1* genomic sequence of SEQ ID No 1, comprising a trait causing mutation, and particularly the mutated genes obtained by the process described above.

A mutated *TBC-1* gene can be defined as a gene encoding either a modified or native TBC-1 protein through a nucleotide sequence which is different from the nucleotide sequence of the *TBC-1* gene found in a majority of trait negative individuals.

Detection of markers or groups of markers associated with a trait

The invention also concerns a method for the detection in an individual of alleles associated with a trait preferably selected from prostate cancer, an early onset of prostate cancer, with a susceptibility to prostate cancer, the level of aggressiveness of prostate cancer tumors, or with the expression of the *TBC-1* gene.

This method comprises the following steps :

- obtaining a nucleic acid sample from the individual to be tested, and
- determining the presence in the sample of an allele of a biallelic marker or of a group of biallelic markers of the *TBC-1* gene which, when taken alone or in combination with another/other biallelic marker/s of the *TBC-1* gene, is indicative of prostate cancer, of an early onset of prostate cancer, of the level of aggressiveness of prostate cancer tumors, of a modified expression of the *TBC-1* gene, of a modified production of the TBC-1 protein, or of the production of a modified TBC-1 protein.

In a preferred embodiment, the biallelic marker comprises the sequences of SEQ ID Nos 7-8.

More particularly, the detection method of the present invention comprises the following steps:

- obtaining a nucleic acid sample from the individual to be tested,
- amplifying a nucleotide sequence of the *TBC-1* gene, particularly of te *TBC-1* genomic sequence of SEQ ID No 1, contained in the sample, and

- detecting the presence in the sample of an allele of a biallelic marker or of a group of biallelic markers of the *TBC-1* gene which, when taken alone or in combination with another/other biallelic markers of the *TBC-1* gene, is indicative of prostate cancer of an early onset of prostate cancer, of the level of aggressiveness of prostate cancer tumors, of a modified expression of the *TBC-1* gene, of a modified production of the TBC-1 protein, or of the production of a modified TBC-1 protein.

In a preferred embodiment, the biallelic marker comprises the sequences of SEQ ID Nos 7-8.

In preferred embodiments of the two detection methods described above, the presence of alleles of one or more biallelic markers of the *TBC-1* gene is determined through microsequencing reactions using microsequencing primers such as those of SEQ ID 11-12. More particularly, it is preferred that the microsequencing primers be bound to a solid support, preferably in the form of arrays of primers attached to appropriate substrates, for example chips, or be used in microfluidic devices. Such arrays are described in further detail in the next section.

Alternatively, the detection method of the present invention can comprise the following steps:

- obtaining a nucleic acid sample from the individual to be tested,
- specifically amplifying a nucleotide sequence of the *TBC-1* gene comprising an allele of biallelic marker or of a group of biallelic markers of a *TBC-1* gene which, when taken alone or in combination with another/other biallelic marker/s of a *TBC-1* gene, is indicative of prostate cancer, of an early onset of prostate cancer, of the level of aggressiveness of prostate cancer tumors, of a modified expression of the *TBC-1* gene, of a modified production of the TBC-1 protein, or of the production of a modified TBC-1 protein, and
- detecting the presence of an amplification product.

In a preferred embodiment, the biallelic marker comprises the sequences of SEQ ID Nos 7-8.

In a further embodiment of the present invention, another detection method comprises the following steps:

- contacting a hybridization probe with a biological sample under conditions which allow hybridization of the probe to the nucleotide sequence of the *TBC-1* gene comprising an allele of a biallelic marker which is indicative of prostate cancer, of an early onset of prostate cancer, of a susceptibility to prostate cancer, of the level of aggressiveness of prostate cancer tumors, of a modified expression of the *TBC-1* gene, of a modified production of the TBC-1 protein, or of the production of a modified TBC-1 protein, and

- detecting the formation of a hybrid comprising the probe and a nucleotide sequence from the biological sample.

In a preferred embodiment, the biallelic marker comprises the sequences of SEQ ID Nos 7-8.

Preferably, the formation of a hybrid is detected through the detection of a signal from a label attached to the probe. More preferably, the signal is amplified prior to being revealed.

In yet another embodiment of the present invention, a further detection method comprises the following steps:

- contacting a hybridization probe immobilized on a solid support with a biological sample under conditions which allow hybridization of the probe to the nucleotide sequence of the *TBC-1* gene comprising an allele of a biallelic marker which is indicative of prostate cancer, of an early onset of prostate cancer, of a susceptibility to prostate cancer, of the level of aggressiveness of prostate cancer tumors, of a modified expression of the *TBC-1* gene, of a modified production of the TBC-1 protein, or of the production of a modified TBC-1 protein, and
- contacting the hybrid thus formed with a labeled hybridization probe, and
- revealing the formation of a hybrid comprising the hybridization probe, a nucleotide sequence from the biological sample and the labeled probe.

In a preferred embodiment, the biallelic marker comprises the sequences of SEQ ID 7-8.

Preferably, the nucleotide sequences of the biological sample are amplified prior to hybridization using one of the primers described herewith.

The invention also specifically relates to a method of determining whether an individual suffering from prostate cancer or susceptible of developing prostate cancer is likely to respond positively to treatment with a selected medicament acting against prostate cancer.

The method comprises the following steps:

- obtaining a DNA sample from the individual to be tested, and
- analyzing said DNA sample to determine whether it comprises alleles of one or more biallelic markers associated with a positive response to treatment with the medicament and/or alleles of one or more biallelic markers associated with a negative response to treatment with the medicament.

Vectors and host cells

A further object of the present invention is a recombinant vector for the expression of an heterologous polynucleotide, wherein said vector comprises a nucleic acid comprising a nucleotide sequence of SEQ ID No 2, or biologically active nucleotide fragments and variants thereof. The heterologous polynucleotide codes either for a desired polypeptide of interest or for a nucleic acid, for example a sense or an antisense DNA molecule.

In a specific embodiment, a recombinant vector of the invention comprises a nucleic acid comprising a polynucleotide encoding a human TBC-1 protein or a fragment or variant thereof.

Another recombinant vector of the invention comprises a nucleic acid comprising a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4, or a regulatory sequence contained therein or the TBC-1 coding sequence contained therein, or fragments or variants thereof. Preferred nucleic acid fragments or variants consist of biologically active fragments or variants thereof.

TBC-1 polypeptides

The invention also concerns a purified or isolated TBC-1 polypeptide encoded by a nucleic acid carrying the *TBC-1* gene, a *TBC-1* cDNA, or a fragment or a variant thereof.

More particularly, the invention also relates to a purified or isolated polypeptide that is encoded by a nucleic acid selected from the group consisting of SEQ ID Nos 3 or 4 or a fragment or a variant thereof.

The invention deals also with a purified or isolated TBC-1 polypeptide that is encoded by a nucleic acid comprising a biallelic marker according to the present invention. In one embodiment, the invention concerns a purified or isolated TBC-1 polypeptide that is encoded by a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 7-8.

More particularly, the invention relates to a purified or isolated TBC-1 polypeptide comprising an aminoacid sequence of SEQ ID No 5 or a fragment or a variant thereof.

Antibodies

The invention also concerns a purified or isolated antibody which is capable of specifically binding to the TBC-1 protein comprising the amino acid sequence of SEQ ID No 5.

The invention also deals with methods and kits for detecting the presence of the polypeptide comprising the amino acid sequence SEQ ID No 5 in a test sample.

The method particularly comprises contacting a test sample suspected of containing the amino acid sequence of SEQ ID No 5 with an antibody of the invention.

The kit comprises an antibody of the invention and preferably means for revealing the formation of an antigen-antibody complex.

Complementary polynucleotides

For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired with its complementary base. Complementary bases are, generally, A and T (or A and U), or C and G.

Diagnostic kits

Another object of the invention consists of diagnostic kits for detecting the presence of at least one copy of a *TBC-1* DNA in a test sample, said kits containing a primer, a pair of primers or a probe according to the invention.

The invention also relates to diagnostic kits useful for determining the presence in a DNA sample of alleles associated with prostate cancer, with an early onset of prostate cancer, with the level of aggressiveness of prostate cancer tumors, with a modified expression of the *TBC-1* gene, with a modified production of the TBC-1 protein, or with the production of a modified TBC-1 protein.

In a first embodiment, the kit comprises primers such as those described above, preferably forward and reverse primers which are used to amplify the *TBC-1* gene, particularly the genomic sequence of SEQ ID No 1, or a fragment thereof. In some embodiments, at least one of the primers is complementary to a nucleotide sequence of the *TBC-1* gene comprising a biallelic marker associated with prostate cancer, with an early onset of prostate cancer, with the level of aggressiveness of prostate cancer tumors, with a modified expression of the *TBC-1* gene, with a

modified production of the TBC-1 protein, or with the production of a modified TBC-1 protein. In one embodiment, the biallelic marker comprises one of the sequences of SEQ ID Nos 7-8.

In a second embodiment, the kit comprises microsequencing primers, wherein at least one of said primers is an oligonucleotide capable of hybridizing, either with the coding or with the non-coding strand, immediately upstream of the polymorphic base of a biallelic marker, preferably the biallelic marker consisting of the group of nucleotide sequences of SEQ ID Nos 7-8, and most preferably the nucleotide sequences of SEQ ID Nos 11-12. Other primers that are part of the invention and which may be included in this kit are primers comprising the nucleotide sequences of SEQ ID Nos 13 and 14.

In a third embodiment, the kit comprises a primer which is complementary to any nucleotide sequence of the *TBC-1* gene, and particularly of the genomic sequence of SEQ ID No 1, and is used to amplify the *TBC-1* gene or a fragment thereof contained in the nucleic acid sample to be tested which includes a polymorphic base of at least one biallelic marker. Preferably, the amplified region includes a polymorphic base of at least one biallelic marker consisting of the nucleotide sequences selected from the group consisting of SEQ ID nos 7-8.

In a fourth embodiment, the kit comprises a hybridization DNA probe, that is or eventually becomes immobilized on a solid support, which is capable of hybridizing with the *TBC-1* gene, particularly with the genomic sequence of SEQ ID No 1, or fragment thereof, preferably which is capable of hybridizing with a region of the *TBC-1* gene which comprises an allele of a biallelic marker associated with prostate cancer, with an early onset of prostate cancer, with a susceptibility to prostate cancer, with the level of aggressiveness of prostate cancer tumors, with a modified expression of the *TBC-1* gene, with a modified production of the TBC-1 protein, or with the production of a modified TBC-1 protein. The techniques for immobilizing a nucleotide primer or probe on a solid support are well-known to the skilled person and include, but are not limited to, the immobilization techniques described in the present application. In a preferred embodiment, the probe is selected from the group consisting of SEQ ID Nos 7-8 or fragment thereof.

The kits of the present invention can also comprise optional elements including appropriate amplification reagents such as DNA polymerases when the kit comprises primers, reagents useful in hybridization reactions and reagents useful to reveal the presence of a hybridization reaction between a labeled hybridization probe and the *TBC-1* gene containing at least one

biallelic marker. In one embodiment, the biallelic marker comprises one of the sequences of SEQ ID Nos 7-8.

Treatment of prostate cancer

The invention also concerns a method for the treatment of prostate cancer comprising the following steps:

- selecting an individual whose DNA comprises an allele of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with a susceptibility to prostate cancer;
- monitoring in said individual the appearance and optionally the development of a tumor in the prostate; and
- administering an effective amount of a medicament acting against prostate cancer to said individual at an appropriate stage of the prostate cancer.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

As used herein, the term "susceptibility to prostate cancer" is intended to designate a strong likelihood for an individual to develop in his or her lifetime a form of prostate cancer. This likelihood is believed to be strongly related to the association established between the biallelic markers of the present invention and prostate cancer or other more specific characteristics which can lead to the development of the prostate cancer such as the modified expression of the *TBC-1* gene, the modified production of the TBC-1 protein or the production of a modified TBC-1 protein.

The term « treatment of prostate cancer » when used herein is intended to designate the administration of substances either for prophylactic or curative purposes. When administered for prophylactic purposes, the treatment is provided in advance of the appearance of biologically or clinically significant cancer symptoms. When administered for curative purposes, the treatment is provided to attenuate the pathological symptoms of prostate cancer, to decrease the size or growth of cancer tumors or metastases or to remove them.

The prophylactic administration of a treatment serves to prevent, attenuate or inhibit the growth of cancer cells.

Another embodiment of the present invention is a method for the treatment of prostate cancer comprising the following steps:

- selecting an individual whose DNA comprises an allele of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene associated with a susceptibility to prostate cancer;
- administering to said individual, preferably as a preventive treatment of prostate cancer, an effective amount of a medicament acting against prostate cancer.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

In a further embodiment, the present invention concerns a method for the treatment of prostate cancer comprising the following steps:

- selecting an individual whose DNA comprises an allele of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with a susceptibility to prostate cancer;
- administering to said individual, as a preventive treatment of prostate cancer, an effective amount of a medicament acting against prostate cancer;
- monitoring in said individual the appearance and optionally the development of a tumor in the prostate; and, if required
- administering an effective amount of a medicament acting against prostate cancer to said individual at an appropriate stage of the prostate cancer.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

The present invention also concerns a method for the treatment of prostate cancer comprising the following steps:

- selecting an individual suffering from a prostate cancer and whose DNA comprises an allele of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with an aggressive form of prostate cancer tumors; and
- administering an effective amount of a medicament acting against prostate cancer to said individual.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8. In particular embodiments, the individual is selected by genotyping one or more biallelic markers of the present invention.

The invention also concerns a method for the treatment of prostate cancer in a selected population of individuals. The method comprises :

- selecting an individual suffering from prostate cancer and whose DNA comprises an allele of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with a positive response to treatment with an effective amount of a medicament acting against prostate cancer, and/or whose DNA does not comprise an allele of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with a negative response to treatment with said medicament; and
- administering at suitable intervals an effective amount of said medicament to said selected individual.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8. In particular embodiments, the individual is selected by genotyping one or more biallelic markers of the present invention.

In the context of the present invention, a "positive response" to a medicament can be defined as comprising a reduction of the symptoms related to the disease or condition to be treated.

In the context of the present invention, a "negative response" to a medicament can be defined as comprising either a lack of positive response to the medicament which does not lead to a symptom reduction or to a side-effect observed following administration of the medicament.

Methods for screening candidate substances or molecules of interest.

Another object of the present invention consists of methods and kits for the screening of candidate substances that interact with the TBC-1 protein described herein as well as methods and kits for the screening of substances that are able to modulate the expression of the *TBC-1* gene.

1. Substances or molecules interacting with the TBC-1 protein according to the invention.

The present invention pertains to methods for screening substances or molecules that interact with the TBC-1 protein or one peptide fragment or variant thereof.

In a first screening method embodiment, such a method for the screening of a candidate substance comprises the following steps :

a)providing a polypeptide comprising the amino acid sequence SEQ ID No 5, or a peptide fragment or a variant thereof;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance;

d) detecting the complexes formed between said polypeptide and said candidate substance.

In one embodiment of the screening method defined above, the complexes formed between the polypeptide and the candidate substance are further incubated in the presence of a polyclonal or a monoclonal antibody that specifically binds to the TBC-1 protein of the invention or to the peptide fragment or variant thereof.

The candidate substance or molecule to be assayed for interacting with the TBC-1 polypeptide may be of diverse nature, including, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides.

In another embodiment of the present screening method, increasing concentrations of a substance competing for binding to the TBC-1 protein with the considered candidate substance is added, simultaneously or prior to the addition of the candidate substance or molecule, when performing step c) of said method. By this technique, the detection and optionally the quantification of the complexes formed between the TBC-1 protein or the peptide fragment or variant thereof and the candidate substance or molecule to be screened allows the one skilled in the art to determine the affinity value of said substance or molecule for said TBC-1 protein or the peptide fragment or variant thereof.

The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a TBC-1 protein having the amino acid sequence of SEQ ID No 5 or a peptide fragment or a variant thereof, and optionally means useful to detect the complex formed between the TBC-1 protein or its peptide fragment or variant and the candidate substance. In a preferred embodiment the detection means consist in monoclonal or polyclonal antibodies directed against the TBC-1 protein or a peptide fragment or a variant thereof.

2. Substances or molecules modulating the expression of the *TBC-1* gene.

The present invention concerns also a method for screening substances or molecules that are able to increase, or in contrast to decrease or even suppress the expression of the *TBC-1* gene. Such a method may allow the one skilled in the art to select substances exerting a regulating effect on the expression level of the *TBC-1* gene and which may be useful for expressing a desired polynucleotide, for example specifically in a prostate tissue.

The expressions "gene expression" or "gene expression level" according to the invention are intended to designate gene activity both at the transcriptional and post-transcriptional level.

The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a recombinant vector that allows the expression of a nucleic acid comprising a nucleotide sequence of SEQ ID No : 2 or alternatively a recombinant cell host containing such a recombinant vector.

Thus, is also part of the present invention a method for screening of a candidate substance or molecule that modulates the expression of the *TBC-1* gene according to the invention, wherein this method comprises the following steps :

a) providing a recombinant cell host containing a nucleic acid, wherein said nucleic acid comprises the nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof, the nucleotide sequence of SEQ ID No 2 or its biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein;

b) obtaining a candidate substance, and

c) determining the ability of the candidate substance to modulate the expression levels of the polynucleotide encoding the detectable protein.

In a preferred embodiment of the above screening method, the nucleic acid comprising the nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof also includes a 5'UTR region of one of the *TBC-1* cDNAs of SEQ ID Nos 3 and 4, or one of their biologically active fragments or variants thereof.

The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise

a) a recombinant vector that allows the expression of a nucleic acid comprising a nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof;

b) a polynucleotide encoding a detectable protein which is operably linked to the nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof.

A second method for the screening of a candidate substance or molecule that modulates the expression of the *TBC-1* gene comprises the following steps :

a) providing a recombinant cell host containing a nucleic acid, wherein said nucleic acid comprises a 5'UTR sequence of one of the *TBC-1* cDNAs of SEQ ID Nos 3 and 4, or one of their biologically active fragments or variants, the 5'UTR sequence or its biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein;

b) obtaining a candidate substance, and

c) determining the ability of the candidate substance to modulate the expression levels of the polynucleotide encoding the detectable protein.

In a preferred embodiment of the screening method described above, the nucleic acid that comprises a nucleotide sequence selected from the group consisting of the 5'UTR sequence of one of the *TBC-1* cDNAs of SEQ ID Nos 3 and 4 or one of their biologically active fragments or variants, includes a promoter sequence, wherein said promoter sequence can be either endogenous, or in contrast exogenous with respect to the *TBC-1* 5'UTR sequences defined therein.

A preferred endogenous promoter sequence consists in a purified or isolated nucleic acid comprising the *TBC-1* regulatory sequence of SEQ ID No 2, or a biologically active fragment or variant thereof.

Among the preferred polynucleotides encoding a detectable protein, there may be cited polynucleotides encoding beta galactosidase, green fluorescent protein (GFP) and chloramphenicol acetyl transferase (CAT).

In another embodiment of a screening method according to the invention, the polynucleotide encoding a detectable protein is selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4, and coding fragments or variants thereof.

The invention also pertains to kits useful for performing the hereinbefore described second screening method. Preferably, such kits comprise

a) a recombinant vector that comprises a nucleic acid including a 5'UTR sequence of one of the *TBC-1* cDNAs of SEQ ID Nos 3 and 4, or one of their biologically active fragments or variants, the 5'UTR sequence or its biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein.

Preferably, the regulatory sequence contained in the recombinant vector described above is located upstream the polynucleotide encoding a detectable protein.

For the design of suitable recombinant vectors useful for performing the screening methods described above, it will be referred to the section of the present specification wherein the preferred recombinant vectors of the invention are detailed.

Variants and fragments

1. Polynucleotides

The invention also relates to variants and fragments of the polynucleotides described herein.

Variants of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

Variants of polynucleotides according to the invention include, without being limited to, nucleotide sequences which are at least 95% identical to a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1-4 or to any polynucleotide fragment of at least 8 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1and 2 , and preferably at least 99% identical, more particularly at least 99.5% identical, and most preferably at least 99.8% identical to a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1-4 or to any polynucleotide fragment of at east 8 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1-2.

Changes in the nucleotide of a variant may be silent, which means that they do not alter the amino acids encoded by the polynucleotide.

However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

In the context of the present invention, particularly preferred embodiments are those in which the polynucleotides encode polypeptides which retain substantially the same biological function or activity as the mature TBC-1 protein.

A polynucleotide fragment is a polynucleotide having a sequence that entirely is the same as part but not all of a given nucleotide sequence, preferably the nucleotide sequence of the *TBC-1* gene, and variants thereof. The fragment can be a portion of the regulatory sequences of the *TBC-1* gene.

Such fragments may be "free-standing", i.e. not part of or fused to other polynucleotides, or they may be comprised within a single larger polynucleotide of which they form a part or region. However, several fragments may be comprised within a single larger polynucleotide.

As representative examples of polynucleotide fragments of the invention, there may be mentioned those which have from about 4, 6, 8, 15, 20, 25, 40, 10 to 20, 10 to 30, 30 to 55, 50 to 100, 75 to 100 or 100 to 200 nucleotides in length.

2. Polypeptides.

The invention also relates to variants, fragments, analogs and derivatives of the polypeptides described herein.

The variant may be 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the TBC-1 protein of the invention is fused with another compound, such as a compound to

increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the TBC-1 protein of the invention under consideration, such as a leader or secretory sequence or a sequence which is employed for purification of this TBC-1 protein or a preprotein sequence. Such variants are deemed to be within the scope of those skilled in the art.

A polypeptide fragment is a polypeptide having a sequence that entirely is the same as part but not all of a given polypeptide sequence, preferably a polypeptide encoded by the *TBC-1* gene and nucleotide variants thereof. Preferred fragments include those regions possessing antigenic properties and which can be used to raise antibodies against the TBC-1 protein.

Such fragments may be "free-standing", i.e. not part of or fused to other polypeptides, or they may be comprised within a single larger polypeptide of which they form a part or region. However, several fragments may be comprised within a single larger polypeptide.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which have from about 5, 6, 7, 8, 9 or 10 to 15, 10 to 20, 15 to 40, or 30 to 55 amino acids long. Preferred are those fragments containing at least one amino acid mutation in the TBC-1 protein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 :

A) Partial *TBC-1* genomic sequence, first cDNA molecule and associated coding sequence (CDS).

B) Partial *TBC-1* genomic sequence, second cDNA molecule and associated coding sequence (CDS).

Figure 2 : An amino acid alignment of a portion of the amino acid sequence of the TBC-1 protein of SEQ ID No 5 with other proteins sharing amino acid homology with TBC-1.

DETAILED DESCRIPTION OF THE INVENTION

The present invention concerns polynucleotides and polypeptides related to the human *TBC-1* gene (also termed "*TBC-1* gene" throughout the present specification) , which is potentially involved in the regulation of the differentiation of various cell types in mammals. A deregulation

or an alteration of *TBC-1* expression, or alternatively an alteration in the amino acid sequence of the TBC-1 protein may be involved in the generation of a pathological state related to cell differentiation in a patient, more particularly to abnormal cell proliferation leading to cancer states, such as prostate cancer.

The identification of genes involved in a particular trait such as prostate cancer susceptibility can be carried out through two main strategies currently used for genetic mapping : linkage analysis and association studies. Linkage analysis requires the study of families with multiple affected individuals and is now useful in the detection of mono- or oligogenic inherited-traits. Conversely, association studies examine the frequency of marker alleles in unrelated trait positive (T+) individuals compared with trait negative (T-) controls, and are generally employed in the detection of polygenic inheritance.

Candidate region on the chromosome 4 (linkage analysis).

In order to localize the prostate cancer gene(s) starting from families, a systematic familial study of genetic link research is carried out using markers of the microsatellite type described at the Genethon laboratory by the Jean Weissenbach team (Dib et al., 1996), the disclosure of which is incorporated herein by reference.

The studies of genetic link or of "linkage" are based on the principle according to which two neighboring sequences on a chromosome do not present (or very rarely present) recombinations by crossing-over during meiosis. To do this, microsatellite DNA sequences (chromosomal markers) constantly co-inherited with the disease studied are searched for in a family having a predisposition for this disease. These DNA sequences organized in the form of a repetition of di-, tri- or tetranucleotides are systematically present along the genome, and thus allow the identification of chromosomal fragments harboring them. More than 5000 microsatellite markers, have been localized with precision on the genome as a result of the first studies on the genetic map carried out at Genethon under the supervision of Jean Weissenbach, and on the physical map (using the "Yeast Artificial Chromosomes"), work conducted by Daniel Cohen at C.E.P.H. and at Genethon (Chumakov et al., 1995). Genetic link analysis calculates the probabilities of recombinations of the target gene with the chromosomal markers used, according to the genealogical tree, the transmission of the disease, and the transmission of the markers. Thus if a particular allele of a given marker is transmitted with the disease more often than chance would have it (recombination level of between 0 and 0.5), it is possible to deduce that the target gene in question is found in the neighborhood of the marker.

Using this technique, it has been possible to localize several genes of genetic predisposition to familial cancers. In order to be able to be included in a genetic link study, the families affected by a hereditary form of the disease must satisfy the "informativeness" criteria: several affected subjects (and whose constitutional DNA is available) per generation, and at best having a large number of siblings.

By linkage analysis, the inventors have identified a candidate region for prostate cancer on chromosome 4. Indeed, the LOD scores at 2 points between the disease and the markers on a total population of approximately fifty families present a value of 2.49 for marker D4S398 which indicates a probable genetic link with this marker. The curve of the variation of the LOD score on a map of 5 markers is centered on D4S398 and the value higher than 3.3 indicates that a gene involved in familial prostate cancer is probably found in the region located between markers D4S2978 and D4S3018, or a space of approximately 9.7 cM.

The inventors have subsequently established a physical map of the candidate region. Namely they have covered the candidate region with a group of cloned and ordered genomic DNA fragments.

Then, biallelic markers located on the human chromosome 4 candidate region described above have been generated by designing suitable primers able to amplify different nucleotide sequences contained in the genomic candidate region using a pool of DNAs from individuals or using a collection of separate DNA samples from individuals. The amplified nucleic acids are then sequenced in order to identify polymorphisms between individuals, and more preferably single nucleotide polymorphisms (SNPs) that define biallelic markers.

As it will be described in detail further in this specification, the inventors have found that the presence of one among the two alleles of a novel particular biallelic marker (namely, marker 99-430-352) located in the chromosome 4 candidate region described hereinbefore is statistically closely related to the occurrence of prostate cancer in individuals harboring this allele of biallelic marker 99-430-352 within their genome, and particularly with sporadic prostate cancer cases.

Based on the linkage analysis and on the results of the association between the biallelic marker 99-430-352 and prostate cancer described above, the chromosome 4 genomic region surrounding this biallelic marker has been suspected by the inventors to contain a genetic determinant involved in prostate cancer, either as a causal determinant of the disease or as a

co-factor involved in the development of the malignancy state leading to prostate cancer. In order to characterize this genetic determinant, the inventors have sequenced the genomic region surrounding biallelic marker 99-430-352. The sequencing of this genomic region have allowed the inventors to identify a novel human gene, the nucleotide sequence of which harbors biallelic marker 99-430-352.

Homologies of the novel human gene translation product with a known murine protein.

Database homology searches have allowed the inventors to determine that the translation product of this novel human gene has significant identity with a murine protein called tbc1. The novel human gene of the invention has thus been called *TBC-1* throughout the present specification. *TBC-1* comprises an open Reading frame that encodes a novel protein, the TBC-1 protein. Based on sequence similarity, an alignment of a portion of the TBC-1 amino acid sequence with the known tbc1 murine protein, it is expected that TBC1 protein may play a role in the cell cycle and in differentiation of various tissues. Indeed, the TBC1 protein contains a 200 amino acid domain called the TBC domain that is homologous to regions in the tre2-oncogene and in the yeast regulators of mitosis BUB2 and cdc16.

The cDNA of the murine *tbc1* gene has been described in US Patent No US 5,700,927 and it encodes a putative protein product of 1141 amino acids. The N-terminus of the murine tbc1 protein contains stretches of cysteines and histidines which may form zinc finger structures in the mature polypeptides. The N-terminus also comprises short stretches of basic amino acids which may be involved in a nuclear localization signal. The TBC domain of the murine tbc1 protein contains several tyrosine residues which are conserved in BUB2 and cdc16. The C-terminus of the murine tbc1 protein contains a long stretch of evenly spaced leucine residues which are susceptible to form a leucine zipper motif.

The murine *tbc1* gene has been shown to be highly expressed in testis and kidney. However, lower levels of expression have also be identified in lung, spleen, brain, and heart. Moreover, murine tbc1 is a nuclear protein which is expressed in a cell- and stage-specific manner.

Studies of murine bone marrow have demonstrated that erythroid cells and megakaryocytes expressed substantial levels of the murine tbc1 protein, but none was detected in mature neutrophils. Similarly, spermatogonia do not express murine *tbc1*, but primary and secondary spermatocytes express abundant *tbc1*. Later in the differentiation of the germ cells, the *tbc1* levels appear to decrease in spermatids and active sperm. The differentiation program of

spermatogonia to spermatocytes therefore involves a significant upregulation of murine *tbc1* expression.

The general distribution of murine *tbc1* is not tissue-specific, but is cell-specific within individual tissues and intimately linked to tissue differentiation. The developmental expression of murine *tbc1*, particularly in hematopoietic and germ cells, suggests that this gene plays a role in the terminal differentiation program of several tissues.

Consequently, an alteration in the expression of the *TBC-1* gene or in the amino acid sequence of the TBC-1 protein leading to an altered biological activity of the latter is likely to cause, directly or indirectly, cell proliferation disorders and thus diseases related to an abnormal cell proliferation such as cancer, particularly prostate cancer.

A. TBC-1 GENOMIC SEQUENCES, cDNAs AND TBC-1 REGULATORY POLYNUCLEOTIDES.

The invention concerns a purified or isolated nucleic acid encoding a TBC-1 polypeptide as well as a nucleic acid complementary thereto and a fragment or variant thereof.

TBC-1 genomic sequences

The inventors have sequenced a portion of the *TBC-1* genomic sequence. A map of the partially sequenced *TBC-1* gene is depicted in the upper line of Figure 1-A) and 1-B). The upper line of Figure 1-A) and 1-B) shows the respective locations of the first three exons of the *TBC-1* gene, designated respectively as Exon 1, Exon 1*bis* and Exon 2. The position of the first nucleotide at the 5'-end of each exon is also indicated, the nucleotide at position 1 being the first nucleotide at the 5'-end of the polynucleotide of SEQ ID No 1.

More precisely, the structural characteristics of the *TBC-1* genomic sequence, as represented in Figure 1, are as follows :

a) a regulatory region located between the nucleotide at position 1 and the nucleotide at position 1999 of SEQ ID No 1.

b) a transcribed region, located between the nucleotide at position 2000 and the nucleotide at position 17589 of SEQ ID No1, this coding region comprising Exon 1, Exon 1*bis* and Exon 2 of the *TBC-1* gene.

Exon 1 starts at the nucleotide in position 2000 and ends at the nucleotide in position 2076 of the nucleotide sequence of SEQ ID

Exon 1*bis* starts at the nucleotide in position 12291 and ends at the nucleotide in position 12372 of the nucleotide sequence of SEQ ID No 1.

Exon 2 starts at the nucleotide in position 12739 and ends at the nucleotide 13248 of the nucleotide sequence of SEQ ID No 1.

The translation start site ATG is located within Exon 2 at nucleotide positions 12832-12835 of the nucleotide sequence of SEQ ID No 1.

The *TBC-1* introns defined hereinafter for the purpose of the present invention are not exactly what is generally understood as "introns" by the one skilled in the art and will consequently be further defined below.

Generally, an intron is defined as a nucleotide sequence that is present both in the genomic DNA and in the unspliced mRNA molecule, and which is absent from the mRNA molecule which has already gone through splicing events. In the case of the *TBC-1* gene, the inventors have found that at least two different spliced mRNA molecules are produced when this gene is transcribed, as it will be described in detail in a further section of the specification. The first spliced mRNA molecule comprises Exons 1 and 2 as shown in Figure 1-A). Thus, the genomic nucleotide sequence comprised between Exon 1 and Exon 2 is an intronic sequence as regards to this first mRNA molecule, despite the fact that this intronic sequence contains Exon 1*bis*. In contrast, Exon 1*bis* is of course an exonic nucleotide sequence as regards to the second *TBC-1* mRNA molecule shown in Figure 1-B).

For the purpose of the present invention and in order to make a clear and unambiguous designation of the different nucleic acids encompassed, it has been postulated that the polynucleotides contained both in the nucleotide sequence of SEQ ID No 1 and in any of the nucleotide sequences of SEQ ID Nos 3 or 4 are considered as exonic sequences. Conversely, the polynucleotides contained in the nucleotide sequence of SEQ ID No 1 and located between Exon 1 and Exon 2, but which are absent both from the nucleotide sequence of SEQ ID No 3 and from the nucleotide sequence of SEQ ID No 4 are considered as intronic sequences.

Consequently, Intron 1 of *TBC-1* starts at the nucleotide in position 2077 and ends at the nucleotide in position 12290 of the nucleotide sequence of SEQ ID No 1. Intron 1*bis* starts at the nucleotide in position 12373 and ends at the nucleotide in position 12738 of the nucleotide sequence of SEQ ID No 1.

The nucleic acids defining the *TBC-1* introns described above, as well as their fragments and variants, may be used as oligonucleotide primers or probes in order to detect the presence of a copy of the *TBC-1* gene in a test sample, or alternatively in order to amplify a target nucleotide sequence within the *TBC-1* intronic sequences.

The present invention pertains to a purified or isolated nucleic acid encoding a human TBC-1 protein, wherein said TBC-1 protein comprises an amino acid sequence of SEQ ID No 5, a nucleotide sequence complementary thereto, a fragment or a variant thereof.

The present invention also concerns a purified or isolated nucleic acid comprising at least 20 consecutive nucleotides of the nucleotide sequence of SEQ ID No 1 or a sequence complementary thereto. Such a nucleic acid may comprise at least 25, 30, 40 or 50 consecutive nucleotides of the nucleotide sequence of SEQ ID No 1.

Another object of the invention relates to a purified or isolated nucleic acid comprising at least 8 consecutive nucleotides of an intronic sequence of the *TBC-1* genomic sequence of SEQ ID No 1, and particularly of Intron 1 and Intron 1*bis* as defined above.

Thus, the invention also deals with A nucleic acid probe or primer comprising at least 8 consecutive nucleotides of a polynucleotide selected from the group consisting of :

a) the nucleotide sequence beginning at the nucleotide in position 2077 and ending at the nucleotide in position 12290 of the nucleotide sequence of SEQ ID No 1 (Intron 1);

b) the nucleotide sequence beginning at the nucleotide in position 12373 and ending at the nucleotide in position 12738 of the nucleotide sequence of SEQ Id No 1 (Intron 1*bis*).

TBC-1 cDNAs

The first three exons of the *TBC-1* gene are represented in the upper line of Figure 1 A) and B):

- Exon 1 spans from the nucleotide at position 2000 to the nucleotide at position 2076 of the nucleotide sequence of the nucleotide sequence of SEQ ID No 1; Exon 1 spans from the

nucleotide in position 1 to the nucleotide in position 76 of the nucleotide sequence of SEQ ID No 3;

- Exon 1*bis* spans from the nucleotide at position 12291 to the nucleotide at position 12372 of the nucleotide sequence of SEQ ID No 1; Exon 1*bis* spans from the nucleotide in position 1 to the nucleotide in position 81 of the nucleotide sequence of SEQ ID No 4;

- Exon 2 spans from the nucleotide at position 12739 to the nucleotide at position 13248 of the nucleotide sequence of SEQ ID No 1; Exon 2 spans from the nucleotide in position 77 to the nucleotide in position 586 of the nucleotide sequence of SEQ ID No 3; Exon 2 spans from the nucleotide in position 82 to the nucleotide in position 591 of the nucleotide sequence of SEQ ID No 4.

The inventors have discovered that the expression of the *TBC-1* gene leads to the production of at least two mRNA molecules, respectively a first and a second *TBC-1* transcription product, as the results of alternative splicing events.

The middle line of Figure 1-A) depicts the main structural features of a purified or isolated nucleic acid consisting of a *TBC-1* cDNA corresponding to a first *TBC-1* mRNA molecule that has been obtained after reverse transcribing a mRNA generated after transcription of the *TBC-1* gene. This mRNA has a nucleotide length of about 4 kilobases.

The first transcription product comprises Exons 1 and 2. This cDNA of SEQ ID No 3 includes a 5'-UTR region, spanning the whole Exon 1 and part of Exon 2. This 5'-UTR region starts from the nucleotide at position 1 and ends at the nucleotide at position 170 of the nucleotide sequence of SEQ ID No 3. The cDNA of SEQ ID No 3 includes a 3'-UTR region starting from the nucleotide at position 3729 and ending at the nucleotide at position 3984 of the nucleotide sequence of SEQ ID No 1. This first transcription product harbors a polyadenylation site located between the nucleotide at position 3942 and the nucleotide at position 3947 of the nucleotide sequence of SEQ ID No 3. The ORF encoding the TBC-1 protein is comprised between the nucleotide in position 171 and the nucleotide in position 3728 of the nucleotide sequence of SEQ ID No 3.

The middle line of Figure 1-B) depicts the main structural features of a purified or isolated nucleic acid consisting of a *TBC-1* cDNA corresponding to a second *TBC-1* mRNA molecule. The 5'-end sequence of this second *TBC-1* mRNA, more particularly the nucleotide sequence comprised between the nucleotide in position 1 and the nucleotide in position 458 of the nucleic

acid of SEQ ID No 4 molecule corresponds to the nucleotide sequence of a 5'-EST that has been obtained from a human pancreas cDNA library and characterized following the teachings of the PCT Application No WO 96/34981. This 5'-EST is also part of the invention and is defined by the nucleotide sequence of SEQ ID No 6.

The second *TBC-1* transcription product comprises Exons 1*bis* and 2. This cDNA of SEQ ID No 4 includes a 5'-UTR region starting from the nucleotide at position 1 and ending at the nucleotide at position 175 of the nucleotide sequence of SEQ ID No 4. This second cDNA also includes a 3'-UTR region starting from the nucleotide at position 3734 and ending at the nucleotide at position 3989 of the nucleotide sequence of SEQ ID No 4. This second transcription product harbors a polyadenylation site located between the nucleotide at position 3947 and the nucleotide at position 3952 of the nucleotide sequence of SEQ ID No 4.The ORF encoding the TBC-1 protein is comprised between the nucleotide in position 176 and the nucleotide in position 3733 of the nucleotide sequence of SEQ ID No 4.

Another object of the invention consists of a purified or isolated nucleic acid comprising a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4 and to nucleic acid fragments thereof.

Preferred nucleic acid fragments of the nucleotide sequences of SEQ ID Nos 3 and 4 consist in polynucleotides comprising their respective Open Reading Frames encoding the TBC-1 protein.

Other preferred nucleic acid fragments of the nucleotide sequences of SEQ ID Nos 3 and 4 consist in polynucleotides comprising at least a part of their respective 5'-UTR or 3'-UTR regions.

The invention also pertains to a purified or isolated nucleic acid having at least a 95% of nucleotide identity with any one of the nucleotide sequences of SEQ ID Nos 3 and 4, or a fragment thereof.

"Percentage of nucleotide identity" for the polynucleotides of the invention is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is

calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Optimal alignment of sequences for comparison may be conducted by computerized implementation of known algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wis., or BlastN and BlastX available from the National Center for Biotechnology Information), or by inspection. The comparison window is set to at least 100 consecutive nucleotides, preferably 200 consecutive nucleotides, more preferably to 500 consecutive nucleotides and most preferably to 1000 consecutive nucleotides of the considered nucleotide sequence of the invention.

The nucleotide differences contained in an homologous nucleic acid as regards to the nucleotide sequences of SEQ ID Nos 3 and 4 are generally randomly distributed throughout the entire nucleic acid. Nevertheless, preferred nucleic acids are those wherein the nucleotide differences as regards to the nucleotide sequence of SEQ ID Nos 3 and 4 are predominantly located outside the coding sequences, and more precisely in the 5'-UTR and the 3'-UTR sequences contained in either nucleotide sequences of SEQ ID Nos 3 and 4.

The invention also relates to a purified or isolated nucleic acid comprising at least 20, preferably 25, and most preferably 30 consecutive nucleotides of any one of the nucleotide sequences of SEQ ID Nos 3 and 4. Such a nucleic acid is notably useful as polynucleotide probe or primer specific for the *TBC-1* gene or the *TBC-1* mRNAs and cDNAs.

Regulatory sequences

As already mentioned hereinbefore, the polynucleotide of SEQ ID No 1 contains regulatory sequences in the non-coding 5'-flanking region that borders the *TBC-1* coding region. A particular 5' regulatory sequence of the *TBC-1* gene comprises the nucleotide sequence of SEQ ID No2. .

The promoter activity of the 5'-regulatory region contained in the nucleotide sequence of SEQ ID No 1 can be assessed as described below.

Genomic sequences lying upstream of the *TBC-1* Exons are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, beta galactosidase, or green fluorescent protein. The sequences upstream of the *TBC-1* coding region are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for increasing transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

Promoter sequences within the upstream genomic DNA may be further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter, individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

Thus, the minimal size of the promoter of the *TBC-1* gene can be determined through the measurement of *TBC-1* expression levels. For this assay, an expression vector comprising decreasing sizes from the promoter generally ranging from 2 kb to 100 bp, with a 3' end which is constant, operably linked to *TBC-1* coding sequence or to a reporter gene is used. Cells, which are preferably prostate cells and more preferably prostate cancer cells, are transfected with this vector and the expression level of the gene is assessed.

The strength and the specificity of the promoter of the *TBC-1* gene can be assessed through the expression levels of the gene operably linked to this promoter in different types of cells and tissues. In one embodiment, the efficacy of the promoter of the *TBC-1* gene is assessed in

normal and cancer cells. In a preferred embodiment, the efficacy of the promoter of the *TBC-1* gene is assessed in normal prostate cells and in prostate cancer cells which can present different degrees of malignancy.

The 3'UTR sequences of the *TBC-1* mRNAs appear to include at least one polyadenylation site.

Polynucleotides carrying the regulatory elements located both at the 5' end and at the 3' end of the *TBC-1* cDNAs may be advantageously used to control the transcriptional and translational activity of an heterologous polynucleotide of interest.

The 5'UTR and 3'UTR regions of a gene are of particular importance in that they often comprise regulatory elements which can play a role in providing appropriate expression levels, particularly through the control of mRNA stability.

A 5' regulatory polynucleotide of the invention may include the 5'-untranslated region (5'-UTR) located between the nucleotide at position 1 and the nucleotide at position 170 of SEQ ID No 3, or a biologically active fragment or variant thereof.

Alternatively, a 5'-regulatory polynucleotide of the invention may include the 5'-untranslated region (5'-UTR) located between the nucleotide at position 1 and the nucleotide at position 175 of SEQ ID No 4, or a biologically active fragment or variant thereof.

A 3' regulatory polynucleotide of the invention may include the 3'-untranslated region (3'-UTR) located between the nucleotide at position 3729 and the nucleotide at position 3984 of SEQ ID No 4, or a biologically active fragment or variant thereof.

Alternatively, a 3' regulatory polynucleotide of the invention may include the 3'-untranslated region (3'-UTR) located between the nucleotide at position 3734 and the nucleotide at position 3989 of SEQ ID No 4, or a biologically active fragment or variant thereof.

Thus, the invention also pertains to a purified or isolated nucleic acid which is selected from the group consisting of :

a) a nucleic acid comprising the nucleotide sequence SEQ ID No 2 ;

b) a nucleic acid comprising a biologically active fragment or variant of the nucleic acid of SEQ ID No 2.

Is also part of the present invention a purified or isolated 5'-UTR nucleic acid comprising a nucleotide sequence located between the nucleotide at position 1 and the nucleotide at position 170 of SEQ ID No 3, or a biologically active fragment or variant thereof.

Is also part of the present invention a purified or isolated 5'-UTR nucleic acid comprising a nucleotide sequence located between the nucleotide at position 1 and the nucleotide at position 175 of SEQ ID No 4, or a biologically active fragment or variant thereof.

Is also part of the invention a purified or isolated 3'-UTR nucleic acid comprising a nucleotide sequence located between the nucleotide at position 3729 and the nucleotide at position 3984 of SEQ ID No 3, or a biologically active fragment or variant thereof;

Is also part of the invention a purified or isolated 3'-UTR nucleic acid comprising a nucleotide sequence located between the nucleotide at position 3734 and the nucleotide at position 3989 of SEQ ID No 4, or a biologically active fragment or variant thereof;

The 5'-UTR purified or isolated nucleic acids described above may be included in the nucleic acid that comprises the nucleotide sequence of SEQ ID No 2 or its biologically active fragments and variants.

Preferred fragments of the nucleic acid of SEQ ID No 2 have a length of about 1000 nucleotides, more particularly of about 400 nucleotides, more preferably of about 200 nucleotides and most preferably about 100 nucleotides.

By a "biologically active fragment or variant" of a *TBC-1* regulatory polynucleotide according to the present invention is intended a polynucleotide comprising or alternatively consisting in a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide in a recombinant cell host.

For the purpose of the invention, a nucleic acid or polynucleotide is "functional" as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and translational regulatory information, and if such sequences are "operatively linked" to nucleotide sequences which encode the desired polypeptide or the desired polynucleotide. An operable

linkage is a linkage in which the regulatory nucleic acid and the DNA sequence sought to be expressed are linked in such a way as to permit gene expression.

More precisely, two DNA molecules (such as a polynucleotide containing a promoter region and a polynucleotide encoding a desired polypeptide or polynucleotide) are said to be "operably linked" if the nature of the linkage between the two polynucleotides does not (1) result in the introduction of a frame-shift mutation or (2) interfere with the ability of the polynucleotide containing the promoter to direct the transcription of the coding polynucleotide. The promoter polynucleotide would be operably linked to a polynucleotide encoding a desired polypeptide or a desired polynucleotide if the promoter is capable of effecting transcription of the polynucleotide of interest.

In order, to identify the relevant biologically active polynucleotide derivatives of SEQ ID No 2, the one skill in the art will refer to the book of Sambrook et al. (Sambrook, J. Fritsch, E. F., and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York) in order to use a recombinant vector carrying a marker gene (i.e. beta galactosidase, chloramphenicol acetyl transferase, etc.) the expression of which will be detected when placed under the control of a biologically active derivative polynucleotide of SEQ ID No 2.

Regulatory polynucleotides of the invention may be prepared from any of the nucleotide sequences of SEQ ID No 1 or SEQ ID Nos 3 and 4 by cleavage using the suitable restriction enzymes, the one skill in the art being guided by the book of Sambrook et al. (1989).

Regulatory polynucleotides may also be prepared by digestion of any of the nucleotide sequences of SEQ ID No 1 or SEQ ID No 4 by an exonuclease enzyme, such as Bal31 (Wabiko et al., 1986, DNA, 5(4):305-314).

These regulatory polynucleotides can also be prepared by chemical synthesis, as described elsewhere in the specification, when the synthesis of oligonucleotide probes or primers is disclosed.

The regulatory polynucleotides according to the invention may be advantageously part of a recombinant expression vector that may be used to express a coding sequence in a desired host cell or host organism. The recombinant expression vectors according to the invention are described elsewhere in the specification.

The invention also encompasses a polynucleotide comprising :

a) a nucleic acid comprising a regulatory nucleotide sequence of SEQ ID No 2, or a biologically active fragment or variant thereof;

b) a polynucleotide encoding a desired polypeptide or nucleic acid, operably linked to the nucleic acid comprising a regulatory nucleotide sequence of SEQ ID No 2, or its biologically active fragment or variant.

c) Optionally, a nucleic acid comprising a 3'-UTR regulatory polynucleotide, preferably a 3'UTR regulatory polynucleotide of the invention.

In a preferred embodiment, a polynucleotide such as disclosed above comprises the nucleic acid of SEQ ID No 2 or a fragment, a variant or a biologically active derivative thereof which is operably linked to the 5'end of the polynucleotide encoding the desired polypeptide or polynucleotide.

In another embodiment, a polynucleotide such as that described above comprises the nucleic acid of said 3'-UTR sequence or a fragment, a variant or a biologically active derivative thereof which is located at the 3' end of the polynucleotide encoding the desired polypeptide or nucleic acid, it being understood that a preferred desired nucleic acid consists of a ribonucleic acid useful as antisense molecule.

The desired polypeptide encoded by the above described nucleic acid may be of various nature or origin, encompassing proteins of prokaryotic or eukaryotic origin. Among the polypeptides expressed under the control of a *TBC-1* regulatory region, it may be cited bacterial, fungal or viral antigens. Are also encompassed eukaryotic proteins such as intracellular proteins, such as "house keeping" proteins, membrane-bound proteins, like receptors, and secreted proteins like the numerous endogenous mediators such as cytokines.

The desired nucleic acid encoded by the above described polynucleotide, usually a RNA molecule, may be complementary to a *TBC-1* coding sequence and thus useful as an antisense polynucleotide.

Such a polynucleotide may be included in a recombinant expression vector in order to express a desired polypeptide or a desired polynucleotide in host cell or in a host organism. Suitable recombinant vectors that contain a polynucleotide such as described hereinbefore are disclosed elsewhere in the specification.

Coding regions

As depicted in Figure 1-A) and 1-B), the *TBC-1* open reading frame is contained in the two *TBC-1* mRNA molecules of about 4 kilobases isolated by the inventors

More precisely, the effective *TBC-1* coding sequence is comprised between the nucleotide at position 171 and the nucleotide at position 3728 of SEQ ID No 3, and between the nucleotide at position 176 and the nucleotide at position 3733 of the nucleotide sequence of SEQ ID No 4.

The invention further provides a purified or isolated nucleic acid comprising a polynucleotide selected from the group consisting of a polynucleotide comprising a nucleic acid sequence located between the nucleotide at position 171 and the nucleotide at position 3728 of SEQ ID No 3, or a variant or fragment thereof or a sequence complementary thereto;

The above disclosed polynucleotide that contains only coding sequences derived from the *TBC-1* ORF may be expressed in a desired host cell or a desired host organism, when said polynucleotide is placed under the control of suitable expression signals. Such a polynucleotide, when placed under the suitable expression signals, may be inserted in a vector for its expression.

B. IDENTIFICATION OF BIALLELIC MARKERS

Biallelic markers generally consist of a single base polymorphism and are defined as genome-derived polynucleotides having between 2 and 100, preferably between 20, 30, or 40 and 60, and more preferably about 45 nucleotides in length, which exhibit biallelic polymorphism at one single base position. Each biallelic marker therefore corresponds to two forms of a polynucleotide sequence included in a gene which, when compared with one another, present a nucleotide modification at one position. Usually, the nucleotide modification involves the substitution of one nucleotide for another (for example C instead of T).

However, this nucleotide modification can also involve an insertion or a deletion of at least one nucleotide, preferably between 1 and 5 nucleotides. The nucleotide modification can also involve the presence of several adjacent single base polymorphisms. This type of nucleotide modification is usually called a "variable motif". Generally, a "variable motif" involves the presence of 2 to 10 adjacent single base polymorphisms. In some instances, series of two or

more single base polymorphisms can be interrupted by single bases which are not polymorphic. This is also globally considered to be a "variable motif".

Generation of biallelic markers

There are two preferred methods through which the biallelic markers of the present invention can be generated. In a first method, DNA samples from unrelated individuals are pooled together, following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms.

One of the major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing which must be carried out. Moreover, this method is sufficiently sensitive so that a biallelic marker obtained therewith usually shows a sufficient degree of informativeness for conducting association studies.

In a second method for generating biallelic markers, the DNA samples are not pooled and are therefore amplified and sequenced individually. The resulting nucleotide sequences obtained are then also analyzed to identify significant polymorphisms.

It will readily be appreciated that when this second method is used, a substantially higher number of DNA amplification reactions must be carried out. It will further be appreciated that including such potentially less informative biallelic markers in association studies to identify potential genetic associations with a trait may allow in some cases the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations. This method is usually preferred when biallelic markers need to be identified in order to perform association studies within candidate genes.

In both methods, the genomic DNA samples from which the biallelic markers of the present invention are generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background, or from familial cases.

The number of individuals from whom DNA samples are obtained can vary substantially, preferably from about 10 to about 1000, preferably from about 50 to about 200 individuals. It is usually preferred to collect DNA samples from at least about 100 individuals in order to have

sufficient polymorphic diversity in a given population to generate as many markers as possible and to generate statistically significant results.

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. The preferred source of genomic DNA used in the context of the present invention is the peripheral venous blood of each donor.

The techniques of DNA extraction are well-known to the skilled technician. Details of a preferred embodiment are provided in Example 2.

DNA samples can be pooled or unpooled for the amplification step. DNA amplification techniques are well-known to those skilled in the art. Amplification techniques that can be used in the context of the present invention include, but are not limited to, the polymerase chain reaction (PCR, RT-PCR), the ligase chain reaction (LCR), and techniques such as the nucleic acid sequence based amplification (NASBA).

The PCR technology is the preferred amplification technique used in the present invention. It has been described in several publications including US Patents 4,683,195, 4,683,202 and 4,965,188, the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press) and White (1997). Each of these publications is incorporated by reference. A typical example of a PCR reaction suitable for the purposes of the present invention is provided in Example 2.

One of the aspects of the present invention is a method for the amplification of a *TBC-1* gene, particularly the genomic sequence of SEQ ID No 1, or a fragment or variant thereof in a test sample, preferably using the PCR technology. The method comprises the steps of contacting a test sample suspected of containing the target *TBC-1* sequence or portion thereof with amplification reaction reagents comprising a pair of amplification primers.

Thus, the present invention also relates to a method for the amplification of a *TBC-1* gene sequence, particularly of a portion of the genomic sequence of SEQ ID No 1, or a fragment or a variant thereof in a test sample, said method comprising the steps of :

a) contacting a test sample suspected of containing the targeted *TBC-1* gene sequence or portion thereof with amplification reaction reagents comprising a pair of amplification primers located on either side of the *TBC-1* region to be amplified, and

b) detecting the amplification products.

In one specific embodiment of the above amplification method, the amplification primers are selected from the group consisting of SEQ ID Nos 9-10 and 13-14.

In another embodiment of the above amplification, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to a region of the *TBC-1* gene, particularly a region of the genomic sequence of SEQ ID No 1.

The primers are more particularly characterized in that they have sufficient complementarity with any sequence of a strand of the *TBC-1* gene close to region to be amplified, for example with a sequence of introns adjacent to exons to amplify.

The primers were defined with OSP software (Hillier & Green, 1991). The length of the first primer can range from 10 to 100 nucleotides, preferably from 10 to 50, 10 to 30 or more preferably 10 to 25 nucleotides. All primers contained a common upstream oligonucleotide tail enabling the easy systematic sequencing of the resulting amplification fragments.

Preferred primers of the invention include the nucleotide sequences of SEQ ID Nos 9-10 and 13-14.

The amplification products generated as described above with the primers of the invention are then sequenced using methods known and available to the skilled technician. Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Following gel image analysis and DNA sequence extraction, sequence data are automatically processed with adequate software to assess sequence quality.

A polymorphism analysis software is used that detects the presence of biallelic sites among individual or pooled amplified fragment sequences. Polymorphism search is based on the presence of superimposed peaks in the electrophoresis pattern. These peaks which present distinct colors correspond to two different nucleotides at the same position on the sequence. The polymorphism has to be detected on both strands for validation.

The biallelic marker 99-430-352 of the present invention is disclosed in Table 2 of Example 4b. Its location on the *TBC-1* gene is indicated as features in SEQ ID No 1. The pair of amplification

primers are listed in the sequence listing as SEQ ID Nos 9-10, these primers allowing the amplification of a nucleic acid containing the polymorphic base that defines this biallelic marker.

In the present invention, the biallelic marker 99-430-352 referred to above is defined by two nucleotide sequences corresponding to oligonucleotides of 47 bases in length comprising at the middle either the major or the minor allele of the polymorphic base, listed as SEQ ID Nos 7-8.

The 99-430-352 biallelic marker is located within Intron 1 of the *TBC-1* gene, and forms part of the present invention.

Biallelic markers in linkage disequilibrium with the preferred markers of the invention

As mentioned before, once an association has been demonstrated between a given biallelic marker and a trait T, the discovery of additional biallelic markers associated to T and in LD with one of the biallelic markers disclosed herein can easily be carried out by the skilled person.

The present invention then also concerns biallelic markers in LD with the specific biallelic markers described above and which are expected to present similar characteristics in terms of their respective association with a given trait.

Measure of linkage disequilibrium between markers

LD among a set of biallelic markers having a heterozygosity rate of ca. 50% can be determined by genotyping between 50 and 1000 unrelated individuals. Genotyping is performed through individual reactions as described above.

LD between any pair of biallelic markers comprising at least one of the biallelic markers of the present invention (M_i,M_j) can be calculated for every allele combination (M_{i1},M_{j1} ; M_{i1},M_{j2} ; M_{i2},M_{j1} and M_{i2},M_{j2}), according to the Piazza formula :

$\Delta M_{ik},M_{jl} = \sqrt{}\theta4 - \sqrt{}(\theta4 + \theta3)(\theta4 + \theta2)$, where :

$\theta4$= - - = frequency of genotypes not having allele k at M_i and not having allele l at M_j

$\theta3$= - + = frequency of genotypes not having allele k at M_i and having allele l at M_j

$\theta2$= + - = frequency of genotypes having allele k at M_i and not having allele l at M_j

The skilled person will readily appreciate that other LD calculation methods can be used without undue experimentation.

Hence, once LD has been demonstrated between a trait and a given biallelic marker, all the biallelic markers shown to be in LD with the given biallelic marker are expected to present similar characteristics in terms of their respective association with a given trait. These additional markers which can be identified and sequenced by the skilled person using the teachings of the present application also fall within the scope of the present invention.

An example of identification of additional biallelic markers associated to a trait based on the previous knowledge of the localization of a first marker associated to a given trait is given below.

Biallelic markers in linkage disequilibrium with a particular marker :Apo E4

The following example relating to the identification of markers in LD with the apoE4 allele is representative of the procedures of the present invention in which markers in LD with a target gene are identified. 3 major isoforms of human apolipoprotein E (apoE2, -E3, and -E4) have been identified by isoelectric focusing and are coded for by 3 alleles (ε 2, 3, and 4) of the Apo E gene.

As originally reported by Strittmatter et al. and by Saunders et al. in 1993, the Apo E ε4 allele is strongly associated with both late-onset familial and sporadic Alzheimer's Disease (AD).

Biallelic markers in LD with the Apo E ε4 allele were identified. This example is illustrative of the general principle that the generation of biallelic markers associated with a trait leads to markers in LD with any biallelic marker already known to be associated with the trait.

An Apo E marker was used to screen the human genomic BAC library. A BAC, which gave a unique hybridization signal on chromosomal region 19q13.2.3 by FISH, was selected for finding biallelic markers as follows.

This BAC contained an insert of 205 kb that was subcloned. Fifty BAC subclones were randomly selected and sequenced. Twenty-five subclone sequences were selected and used to design twenty-five couples of PCR primers that allowed amplicons of approximately 500 bp to be generated. These PCR primers were then used to amplify the corresponding genomic sequences in a pool of DNA from 100 individuals (French origin, blood donors) as already described. Amplification products from pooled DNA were sequenced and analyzed for the presence of biallelic polymorphisms using the software described herein. Five amplicons were

shown to contain a polymorphic base in the pool of 100 individuals, and therefore these polymorphisms (99-366/274; 99-344/439; 99-365/344; 99-359/308; 99-355/219) were selected as the random biallelic markers in the vicinity of the Apo E gene.

An additional couple of primers was designed that allowed amplification of the genomic fragment carrying the already known polymorphism of Apo E, (99-2452/54 C/T).

An association study was then performed. As expected, there was a clear association between Alzheimer disease (AD) and the known Apo E4 polymorphism (biallelic marker 99-2452/54), the C allele frequency being increased in 26 % in the AD case population studied compared to the AD control population analyzed (pvalue of this difference = $2x10^{-21}$).

In addition, the association study with the random markers generated in the variety of the Apo E gene showed that the biallelic marker 99-365/344 C/T is also associated to AD, the T allele frequency being increased of 17 % in the AD case population respect to the AD control population under study (pvalue of this allele frequency difference = $7x10^{-10}$). Thus individuals who possess a T allele at the biallelic marker 99-365/344 are at risk of developing AD.

Among the biallelic markers generated in the Apo E region, 99-365/344 is in LD with the previously known Apo E4 marker 99-2452/54. The LD is detected in a control population (LD value = 0.08) and is clearly increased in the AD case population (LD = 0.21). Hence the generated biallelic marker which are associated with Alzheimer's disease, namely the biallelic marker 99-365, is in LD with the biallelic marker 99-2452 already known to be associated with this disease.

C- GENOTYPING OF BIALLELIC MARKERS

The polymorphisms identified above can be further confirmed and their respective frequencies can be determined through various methods using the previously described primers and probes. These methods can also be useful for genotyping either new populations in association studies or individuals in the context of detection of alleles of biallelic markers which are known to be associated with a given trait. Those skilled in the art should note that the methods described below can be equally performed on individual or pooled DNA samples.

Once a given polymorphic site has been found and characterized as a biallelic marker as described above, several methods can be used in order to determine the specific allele carried by an individual at the given polymorphic base.

The identification of biallelic markers described previously allows the design of appropriate oligonucleotides, which can be used as probes and primers, to amplify a *TBC-1* gene containing the polymorphic site of interest and for the detection of such polymorphisms.

Amplification

Most genotyping methods require the previous amplification of the DNA region carrying the polymorphic site of interest. Amplification can be performed using the same primers already detailed or alternative second primers.

The invention also concerns alternative second DNA primers, preferably in the form of primer pairs characterized in that they preferably comprise more than 8 nucleotides, preferably between 8 and 100 nucleotides, more preferably between 8 and 50 nucleotides, further more preferably between 8 and 30 nucleotides and in that they are sufficiently complementary with a region of a *TBC-1* gene to hybridize therewith. In some embodiments, the primer pair is adapted for amplifying a sequence containing the polymorphic base of one of the sequences of SEQ ID Nos 7-8.

For amplification and sequencing, the pairs of primers are sufficiently complementary with a region of a *TBC-1* gene, particularly of the genomic sequence of SEQ ID No 1, located at less than 500 pb, preferably at less than 100 pb, and more preferably at less than 50 pb of a polymorphic site corresponding to one of the markers of the present invention.

One of the techniques that can be applied for the amplification of a polymorphic *TBC-1* gene or fragments thereof in a sample using the second primers of the invention can be selected from the techniques described above for the amplification of the *TBC-1* gene.

These second primers can be used, for example, for specific amplification experiments. In these experiments, at least one primer is sufficiently complementary with a region of a *TBC-1* gene comprising the considered polymorphic site corresponding to one of the markers of the present invention to hybridize therewith that said primer is able to initiate the specific amplification of one

allele of the biallelic marker. In some embodiments, the biallelic markers are comprised in one of sequences of SEQ ID Nos 7-8.

Sequencing

The amplification products generated above with the primers of the invention can be sequenced using methods known and available to the skilled technician. Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. A sequence analysis can allow the identification of the base present at the polymorphic site.

It is to be noted that amplification primers such as those of SEQ ID Nos 9-10 or 13-14 can be immobilized on an appropriate solid support prior to amplification. Various techniques of solid-phase immobilization of nucleotide sequences are disclosed below. a preferred technique is described in PCT application WO 96/13609 incorporated herein by reference.

Microsequencing

Polymorphism analyses on pools or selected individuals of a given population can be carried out by conducting microsequencing reactions on candidate regions comprised in amplified fragments obtained by PCR performed on DNA or RNA samples taken from these individuals.

To do so, DNA samples are subjected to PCR amplification of the candidate regions under conditions similar to those described above. These amplification products are then subjected to automated microsequencing reactions using ddNTPs (specific fluorescence for each ddNTP) and appropriate oligonucleotide microsequencing primers which can hybridize just upstream of the polymorphic base of interest. Once specifically extended at the 3' end by a DNA polymerase using a complementary fluorescent dideoxynucleotide analog (thermal cycling), the primer is precipitated to remove the unincorporated fluorescent ddNTPs. The reaction products in which fluorescent ddNTPs have been incorporated are then analyzed by electrophoresis on ABI 377 sequencing machines.

Microsequencing primers hybridize upstream of the polymorphic base to be genotyped, either with the coding or with the non-coding strand. Preferably, the 3' end of the microsequencing primer is immediately upstream of the polymorphic base of the biallelic marker being genotyped, such that upon extension of the primer, the polymorphic base is the first base incorporated. The microsequencing primers may be oligonucleotides of 10, 15, 20 or more bases in length.

In a preferred embodiments, the microsequencing primers are those indicated as features within the sequence listings corresponding to markers of SEQ ID Nos 7-8.

An example of a typical microsequencing procedure that can be used in the context of the present invention is provided in example 5. It is to be understood that certain parameters of this procedure such as the electrophoresis method or the labeling of ddNTPs could be modified by the skilled person without substantially modifying its result.

As a further alternative to the process described above, several solid phase microsequencing reactions have been developed. The basic microsequencing protocol is the same as described previously, except that either the oligonucleotide microsequencing primers or the PCR-amplified products of the DNA fragment of interest are immobilized. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles.

In such solid phase microsequencing reactions, incorporated ddNTPs can either be radiolabeled (see Syvänen, 1994, incorporated herein by reference) or linked to fluorescein (see Livak & Hainer, 1994, incorporated herein by reference). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate).

Other possible of reporter-detection couples include :
- ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (see Harju et al., 1993, incorporated herein by reference)
- biotinylated ddNTP and horseradish peroxidase-conjugated streptavidin with o-phenylenediamine as a substrate (see WO 92/15712, incorporated herein by reference).

A diagnosis kit based on fluorescein-linked ddNTP with antifluorescein antibody conjugated with alkaline phosphatase is commercialized under the name PRONTO by GamidaGen Ltd.

As yet another alternative microsequencing procedure, Nyren et al. (1993) presented a concept of solid-phase DNA sequencing that relies on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA). The PCR-amplified products are biotinylated and immobilized on beads. The microsequencing primer is annealed

and four aliquots of this mixture are separately incubated with DNA polymerase and one of the four different ddNTPs. After the reaction, the resulting fragments are washed and used as substrates in a primer extension reaction with all four dNTPs present. The progress of the DNA-directed polymerization reactions are monitored with the ELIDA. Incorporation of a ddNTP in the first reaction prevents the formation of pyrophosphate during the subsequent dNTP reaction. In contrast, no ddNTP incorporation in the first reaction gives extensive pyrophosphate release during the dNTP reaction and this leads to generation of light throughout the ELIDA reactions. From the ELIDA results, the first base after the primer is easily deduced.

Probes and hybridization

The invention also relates to a group of probes characterized in that they preferably comprise between 8 and 50 nucleotides, and in that they are sufficiently complementary to a polymorphic sequence defined by a biallelic marker located in the genomic sequence of *TBC-1* to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation.

The length of these probes can range from 8, 10, 15, 20, or 30 to 100 nucleotides, preferably from 8 to 50, more preferably from 40 to 50 nucleotides. Particularly preferred probes range in lenght between 40 and 50 nucleotides, for example 47 nucleotides in length. they include a centrally located (for example at position 24) nucleotide complementary to a polymorphic site of the *TBC-1* gene, preferably a polymorphic site corresponding to one of the biallelic markers of the present invention, such as the polymorphic base in the sequences of SEQ ID Nos 7-8, and a 20 to 25 (for example 22) nucleotide sequence spanning on each side of the central nucleotide and substantially complementary to the nucleotide sequences of the *TBC-1* gene spanning on each side of the polymorphic site.

The probes are generally labeled with a radioactive element (^{32}P, ^{35}S, ^{3}H, ^{125}I) or by a non-isotopic molecule (for example, biotin, acetylaminofluorene, digoxigenin, 5-bromodesoxyuridin, fluorescein).

Examples of non-radioactive labeling of nucleic acid fragments are described in the French patent N° FR-7810975 or by Urdea et al. (Urdea M.S., 1988, Nucleic Acids Research, Vol. 11: 4937-4957) or Sanchez-Pescador et al., (Sanchez-Pescador R., 1988, J. Clin. Microbiol., Vol. 26(10):1934-1938).

Advantageously, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al. in 1991 or in the European patent N° EP-0225,807 (Chiron).

The probes are preferably directly labeled such as with isotopes, reporter molecules or fluorescent labels or indirectly labeled such as with biotin to which a streptavidin complex may later bind. Probe labeling techniques are well-known to the skilled technician. By assaying the presence of the probe, one can detect the presence or absence of the targeted DNA sequence in a given sample. The same labels can be used with primers.

The probes of the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA or Northern hybridization to mRNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in the *TBC-1* gene or mRNA using other techniques. The probes are complementary to the *TBC-1* gene coding sequences, to introns, and to regulatory sequences such as promoter, and most preferably to sequences comprised in the genomic sequence of SEQ ID No 1.

Any of the primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The "solid phase" is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples.

Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A "solid phase", as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid phase can be chosen for its intrinsic ability to attract and immobilize the capture reagent.

Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance

that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent.

As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid phase and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid phase material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes and other configurations known to those of ordinary skill in the art.

Polymorphisms can be analyzed and the frequency of corresponding alleles quantified through hybridization reactions on amplified *TBC-1* sequences. The amplification reaction can be carried out as described previously. The hybridization probes which can be conveniently used in such reactions preferably include the probes defined above as being sufficiently complementary to a polymorphic site defined by one of the biallelic markers located in the genomic sequence of *TBC-1* to hybridize thereto and sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation.

The amplification reagents, detection probes and test sample are preferably placed under amplification conditions whereby, in the presence of the target *TBC-1* sequence, copies of the target *TBC-1* sequence (an amplicon) are produced. The probes can also be incorporated into the sample after the amplification reaction has taken place. In the usual case, the amplicon is double stranded because primers are provided to amplify a target *TBC-1* sequence and its complementary strand. Upon formation of the single stranded amplicon members, the mixture is cooled to allow the formation of complexes between the probes and single stranded amplicon members.

After the probe/amplicon member hybrids are formed, they are detected. Standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes. Preferably, the hybrids can be bound to a solid phase reagent by virtue of a capture label and detected by virtue of a detection label. In cases where the detection label is directly detectable, the presence of the hybrids on the solid phase can be detected by causing the label to produce a detectable signal, if necessary, and detecting the signal. In cases where

the label is not directly detectable, the captured hybrids can be contacted with a conjugate, which generally comprises a binding member attached to a directly detectable label. The conjugate becomes bound to the complexes and the conjugates presence on the complexes can be detected with the directly detectable label. Thus, the presence of the hybrids on the solid phase reagent can be determined. Those skilled in the art will recognize that wash steps may be employed to wash away unhybridized amplicon or probe as well as unbound conjugate.

While the amplification primers initiate amplification of the target *TBC-1* sequence, the detection (or hybridization) probe is not involved in amplification. Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702; morpholino analogs which are described in U.S. Patents Numbered 5,185,444, 5,034,506, and 5,142,047; and the like. Depending upon the type of label carried by the probe, the probe is employed to capture or detect the amplicon generated by the amplification reaction. The probe is not involved in amplification of the target sequence and therefore may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified. U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993 describes modifications which can be used render a probe non-extendable.

Accordingly, the ratio of primers to probes is not important. Thus, either the probes or primers can be added to the reaction mixture in excess whereby the concentration of one would be greater than the concentration of the other. Alternatively, primers and probes can be employed at in equivalent concentrations. Preferably, however, the primers are added to the reaction mixture in excess of the probes. Thus, primer to probe ratios of, for example, 5:1 and 20:1 are preferred.

Consequently, the invention also deals with a method for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 1-4 and 7-8 in a sample, said method comprising the following steps of :

a) bringing into contact a nucleic acid probe or a plurality of nucleic acid probes which can hybridize to a nucleotide sequence included in one of the nucleic acids of SEQ ID Nos 1-4 and 7-8 and the sample to be assayed.
b) detecting the hybrid complex formed between the probe or the plurality of probes and the nucleic acid in the sample.

In a first preferred embodiment of the above method, said nucleic acid probe or the plurality of nucleic acid probes is selected from the group consisting of SEQ ID Nos 9-10 and 11-12.

In a second preferred embodiment of this detection method, said nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule.

In a third preferred embodiment of said method, said nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate.

The invention further concerns a kit for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 1-4 and 11-12 in a sample, said kit comprising :
a) a nucleic acid probe or a plurality of nucleic acid probes which can hybridize to a nucleotide sequence included in one of the nucleic acids of SEQ ID Nos 1-4 and 7-8;
b) optionally, the reagents necessary for performing the hybridization reaction.

The nucleic acid probe or the plurality of nucleic acid probes that are included in the detection kit described above may be selected from the group consisting of SEQ ID Nos 9-10 and 11-12.

In a first preferred embodiment of the detection kit, said nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule.

In a second preferred embodiment of the detection kit, said nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate.

a) **DNA chips technology and hybridization**

DNA chips result from the adaptation of computer chips to biology. They allow the integration of micro-biochemical processes (such as DNA hybridization), systems of signal detection (such as

fluorescence) and data processing into a single system which can be used to obtain information on polymorphism.

Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The immobilization of arrays of DNA probes on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854 and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, the disclosure of which are incorporated herein by reference, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques.

In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the probe arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256, the disclosure of which are incorporated herein by reference.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (see Hacia et al., 1996 ; Shoemaker et al., 1996 ; Kozal et al., 1996, incorporated herein by reference).

At least, three companies propose chips able to detect biallelic polymorphisms: Affymetrix (GeneChip), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

One of the limitations encountered when using DNA chip technology is that hybridization of nucleic acids with the probes attached to the chip in arrays is not simply a solution-phase reaction. A possible improvement consists in using polyacrylamide gel pads isolated from one

another by hydrophobic regions in which the DNA probes are covalently linked to an acrylamide matrix.

For the detection of polymorphisms, probes which contain at least a portion of one of the biallelic markers of the present invention, such as the biallelic markers of SEQ ID Nos. 7-8, are synthesized either *in situ* or by conventional synthesis and immobilized on an appropriate chip using methods known to the skilled technician. The solid surface of the chip is often made of silicon or glass but it can be a polymeric membrane. Thus, in some embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of SEQ ID Nos. 7-8 and 11-12, or the sequences complementary thereto, or a fragment thereof at least 15 consecutive nucleotides. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more sequences selected from the group consisting of SEQ ID Nos. 7-8 and 11-12, or the sequences complementary thereto, or a fragment thereof at least 15 consecutive nucleotides.

The nucleic acid sample which includes the candidate region to be analyzed is isolated, amplified and labeled with a reporter group. This reporter group can be a fluorescent group such as phycoerythrin. The labeled nucleic acid is then incubated with the probes immobilized on the chip using a fluidics station. For example, Manz et al. (1993, the disclosure of which is incorporated herein by reference) describe the fabrication of fluidics devices and particularly microcapillary devices, in silicon and glass substrates.

After the reaction is completed, the chip is inserted into a scanner and patterns of hybridization are detected. The hybridization data is collected, as a signal emitted from the reporter groups already incorporated into the nucleic acid, which is now bound to the probes attached to the chip. Probes that perfectly match a sequence of the nucleic acid sample generally produce stronger signals than those that have mismatches. Since the sequence and position of each probe immobilized on the chip is known, the identity of the nucleic acid hybridized to a given probe can be determined.

For single-nucleotide polymorphism analyses, sets of four oligonucleotide probes (one for each base type), preferably sets of two oligonucleotide probes (one for each base type of the biallelic marker) are generally designed that span each position of a portion of the candidate region found in the nucleic acid sample, differing only in the identity of the polymorphic base. The relative intensity of hybridization to each series of probes at a particular location allows the

identification of the base corresponding to the polymorphic base of the probe. Since biallelic polymorphism detection involves identifying single-base mismatches on the nucleic acid sample, greater hybridization stringencies are required (at lower salt concentration and higher temperature over shorter time periods).

The use of direct electric field control improves the determination of single base mutations (Nanogen). A positive field increases the transport rate of negatively charged nucleic acids and results in a 10-fold increase of the hybridization rates. Using this technique, single base pair mismatches are detected in less than 15 sec (see Sosnowski et al., 1997, the disclosure of which is incorporated herein by reference).

The invention is also directed to an array of nucleic acid sequences, said array comprising at least one of the sequences selected from the group consisting of SEQ ID Nos 7-8, 9-10 11-12 or the sequences complementary thereto or a fragment thereof of at least 8 consecutive nucleotides thereof.

In a preferred embodiment of the array described above, this array includes therein at least two of the sequences selected from the group consisting of SEQ ID Nos 7-8, 9-10 and 11-12 or the sequences complementary thereto or a fragment of at least 8 consecutive nucleotides thereof.

b) Integrated microsequencing and capillary electrophoresis chips

Another technique which can be used to analyze polymorphisms includes multicomponent integrated systems which miniaturize and compartmentalize processes such as restriction enzyme, PCR, and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136, the disclosure of which is incorporated herein by reference, which concerns the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts. Regulating or varying the voltage controls the liquid flow at intersections between the micro-machined channels and changes the liquid flow rate for pumping across different sections of the microchip.

In the case of biallelic marker analyses, the micro-chip integrates nucleic acid amplification, a microsequencing reaction (such as the one described above), capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

In a first step, the DNA samples are amplified, preferably by PCR. Then, the amplification products are subjected to automated microsequencing reactions using ddNTPs (specific fluorescence for each ddNTP) and the appropriate oligonucleotide microsequencing primers which hybridize just upstream of the targeted polymorphic base. Once the extension at the 3' end is completed, the primers are separated from the unincorporated fluorescent ddNTPs by capillary electrophoresis. The separation medium used in capillary electrophoresis can for example be polyacrylamide, polyethyleneglycol or dextran. The incorporated ddNTPs in the single-nucleotide primer extension products are identified by fluorescence detection. This micro-chip can be used to process at least 96 to 384 samples in parallel. It can use the usual four color laser induced fluorescence detection of the ddNTPs.

Thus, the chips may comprise an array comprising at least one of the sequences selected from the group consisting of SEQ ID Nos. 7-8, or the sequences complementary thereto or a fragment thereof of at least 8 consecutive nucleotides for determining whether a sample contains one or more alleles of the biallelic markers of SEQ ID Nos. 7-8. The chips may also comprise an array comprising at least one of the sequences selected from the group consisting of SEQ ID Nos. 9-10 or the sequences complementary thereto or a fragment thereof of at least 8 consecutive nucleotides for amplifying one or more alleles of the biallelic markers of SEQ ID Nos. 7-8. In other embodiments, the chips may also comprise an array comprising at least one of the sequences selected from the group consisting of SEQ ID Nos. 9-10 or the sequences complementary thereto or a fragment thereof of at least 8 consecutive nucleotides for conducting microsequencing analyses to determine whether a sample contains one or more alleles of the biallelic markers of SEQ ID Nos. 7-8. In still further embodiments, the chip may comprises an array comprising at least one of the sequences selecting from the group consisting of SEQ ID Nos. 11-12 or the sequences complementary thereto or a fragment thereof of at least 8 consecutive nucleotides for determining whether a sample contains one or more alleles of the biallelic markers of SEQ ID Nos. 7-8.

c) **Other methods**

Other methods well-known by those skilled in the art can also be used to detect biallelic polymorphisms such as single strand conformational polymorphism analysis (SSCP) described in Orita et al. (1989), the disclosure of which is incorporated herein by reference, denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield, V.C. et al. (1991), White et al. (1992), Grompe et al. (1989), and Grompe (1993), the disclosures of which are incorporated herein by reference. These methods are well known to those skilled in the art.

The biallelic markers according to the present invention may be used in methods for the identification and characterization of an association between alleles for one or several biallelic markers of the sequence of the *TBC-1* gene and a trait.

The identified polymorphisms, and consequently the biallelic markers of the invention, may be used in methods for the detection in an individual of *TBC-1* alleles associated with a trait, more particularly a trait related to a cell differentiation or abnormal cell proliferation disorders, and most particularly a trait related to cancer diseases, specifically prostate cancer.

D- ASSOCIATION STUDIES

As mentioned before, the identification of genes involved in polygenic traits such as prostate cancer susceptibility or response to treatment against prostate cancer, can be carried out by performing association analyses. As it will be described below, the purpose of association studies is to compare the frequencies of biallelic markers in trait positive and trait negative populations. If a trait-causing allele (TCA) is associated to a trait T, then the frequency of this particular TCA as well as that of all alleles in LD with that particular TCA, will be significantly higher in trait + than in trait – individuals.

The general strategy to perform association studies using biallelic markers derived from a candidate gene is to scan two groups of individuals (trait + and trait – control individuals which are characterized by a well defined phenotype as described below) in order to measure and statistically compare the allele frequencies of such biallelic markers in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analyzed biallelic markers, one can assume that : either the associated allele is directly

responsible for causing the trait (associated allele is the TCA), or the associated allele is in LD with the TCA. The specific characteristics of the associated allele with respect to the candidate gene function usually gives further insight into the relationship between the associated allele and the trait (causal or in LD). If the evidence indicates that the associated allele within the candidate gene is most probably not the TCA but is in LD with the real TCA, then the TCA can be found by sequencing the vicinity of the associated marker.

Collection of DNA samples from trait positive (trait +) and trait negative (trait – individuals (inclusion criteria)

In order to perform efficient and significant association studies such as those described herein, the trait under study should preferably follow a bimodal distribution in the population under study, presenting two clear non-overlapping phenotypes, trait + and trait -.

Nevertheless, even in the absence of such a bimodal distribution (as may in fact be the case for more complex genetic traits), any genetic trait may still be analyzed by the association method proposed here by carefully selecting the individuals to be included in the trait + and trait – phenotypic groups. The selection procedure involves to select individuals at opposite ends of the non-bimodal phenotype spectra of the trait under study, so as to include in these trait + and trait – populations individuals which clearly represent extreme, preferably non-overlapping phenotypes.

The definition of the inclusion criteria for the trait + and trait – populations is an important aspect of the present invention. The selection of drastically different but relatively uniform phenotypes enables efficient comparisons in association studies and the possible detection of marked differences at the genetic level, provided that the sample sizes of the populations under study are significant enough.

Generally, trait + and trait – populations to be included in association studies such as proposed in the present invention consist of phenotypically homogenous populations of individuals each representing 100% of the corresponding trait if the trait distribution is bimodal.

A first group of between 50 and 300 trait + individuals, preferably about 100 individuals, are recruited according to clinical inclusion criteria based on either 1°) affection by prostate cancer, 2°) evidence of aggressiveness of prostate cancer tumors.

In each case, a similar number of trait – individuals, preferably more than 100 individuals, are included in such studies who are preferably both ethnically- and age-matched to the trait positive cases. They are checked for the absence of the clinical criteria defined above. Both trait + and trait - individuals should correspond to unrelated cases.

In one of the preferred embodiments of the present invention, association studies are carried out on the basis of a presence (trait+) or absence (trait-) of prostate cancer. Trait- group of individuals is described in details in Example 6. Several groups of trait+ individuals were used : (1) independent individuals consisting of both patients diagnosed before the age of 65 years old and patients diagnosed after the age of 65 years old, these two later groups being also tested separately; (2) groups of unrelated individuals consisting of prostate cancer sporadic cases and (3) groups of unrelated individuals consisting of familial prostate cancer cases consisting of both patients diagnosed before the age of 65 years old and patients diagnosed after the age of 65 years old, these two later groups being also tested separately. A specific protocol for the collection of DNA samples from trait + and trait – individuals is described in Example 6.

In order to have as much certainty as possible on the absence of prostate cancer in trait – individuals, it is preferred to conduct a PSA dosage analysis on this population. Several commercial assays can be used (WO 96/21042, herein by reference). In one preferred embodiment, a Hybritech assay is used and trait – individuals must have a level of PSA less than 2.8 ng/ml of serum in order to be selected as such. In a preferred embodiment, the Yang assay is used and trait – individuals must have a level of PSA of less than 4 ng/ml of serum in order to be included in the population under study.

Genotyping of trait + and trait - individuals

Allelic frequencies of the biallelic markers in each of the above described population can be determined using one of the methods described above under the heading "Genotyping of biallelic markers". Analyses are preferably performed on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual in similar conditions as those described above for the generation of biallelic markers.

In a preferred embodiment, amplified DNA samples are subjected to automated microsequencing reactions using fluorescent ddNTPs (specific fluorescence for each ddNTP) and the appropriate microsequencing oligonucleotides which hybridize just upstream of the polymorphic base.

Genotyping is further described in Example 7.

Association studies

In one preferred embodiment of the invention, a correlation was found between the 99-430-352 biallelic marker of the TBC-1 gene and prostate cancer, results of the association study are further described in details in example 8.

Similar association studies can also be carried out with other biallelic markers within the scope of the invention, preferably with biallelic markers in LD with the markers associated with prostate cancer as described above, including the biallelic markers of SEQ ID Nos 7-8.

Similar associations studies can be carried out by the skilled technician using the biallelic markers of the invention defined above, with different trait + and trait - populations. Suitable further examples of association studies using biallelic markers of the *TBC-1* gene, including the biallelic markers of SEQ ID Nos 7-8, involve studies on the following populations:

- a trait + population suffering from a cancer and a healthy unaffected population, or
- a trait + population suffering from prostate cancer treated with agents acting against prostate cancer and suffering from side-effects resulting from this treatment and an trait - population suffering from prostate cancer treated with same agents without any substantial side-effects, or
- a trait + population suffering from prostate cancer treated with agents acting against prostate cancer showing a beneficial response and a trait – population suffering from prostate cancer treated with same agents without any beneficial response, or
- a trait + population suffering from prostate cancer presenting highly aggressive prostate cancer tumors and a trait – population suffering from prostate cancer with prostate cancer tumors devoid of aggressiveness.

E- IDENTIFICATION OF A TRAIT CAUSING MUTATION IN THE *TBC-1* GENE

Mutations in the *TBC-1* gene which are responsible for a detectable phenotype may be identified by comparing the sequences of the *TBC-1* genes from trait + and trait - individuals. Preferably, trait + individuals to be sequenced carry a single marker allele or a haplotype shown to be associated to the trait and trait – individuals to be sequenced do not carry such allele or haplotype associated to the trait. The detectable phenotype may comprise prostate cancer, a response to or side effects related to a prophylactic or curative agent acting against prostate cancer, the aggressiveness of prostate cancer tumors, expression of the *TBC-1* gene, a

modified or forthcoming production of the TBC-1 protein, or the production of a modified TBC-1 protein. The mutations may comprise point mutations, deletions, or insertions in the *TBC-1* gene. The mutations may lie within the coding sequence for the TBC-1 protein or within regulatory regions in the *TBC-1* gene, such as its promoter.

The method used to detect such mutations generally comprises the following steps :

- amplification of a region of the *TBC-1* gene comprising a biallelic marker or a group of biallelic markers associated to the considered trait from DNA samples of trait + patients and trait - controls;
- sequencing of the amplified region;
- comparison of DNA sequences from trait + patients and trait - controls; and
- determination of mutations specific to trait + patients.

Oligonucleotide primers are constructed as described previously to amplify the sequences of each of the exons, introns, the promoter region and the regulatory regions of the *TBC-1* gene. The oligonucleotide primers may comprise at least 10, preferably 15, more preferably 20, and even more preferably 25 or more consecutive nucleotides of the *TBC-1* genomic DNA or the *TBC-1* cDNA or the sequences complementary thereto.

Each primer pair is used to amplify the exon, promoter region, introns or regulatory regions, from which it is derived. Amplification is carried out on genomic DNA samples from trait + patients and trait - controls, preferably using the PCR conditions described in the examples. Amplification products from the genomic PCRs are then subjected to sequencing, preferably through automated dideoxy terminator sequencing reactions and electrophoresed, preferably on ABI 377 sequencers. Following gel image analysis and DNA sequence extraction, ABI sequence data are automatically analyzed to detect the presence of sequence variations among trait + and trait - individuals. Sequences are verified by determining the sequences of both DNA strands for each individual.

Candidate polymorphisms suspected of being responsible for the detectable phenotype, are then verified by screening a larger population of trait + and trait - individuals using polymorphism analysis techniques such as the techniques described above. Polymorphisms which exhibit a statistically significant correlation with the detectable phenotype are deemed responsible for the detectable phenotype.

In another embodiment, the mutant *TBC-1* allele which causes a detectable phenotype can be isolated by obtaining a nucleic acid sample such as a genomic library or a cDNA library from an individual expressing the detectable phenotype. The nucleic acid sample can be contacted with one or more probes lying in the region of the *TBC-1* gene where the associated biallelic marker or group of biallelic markers or with PCR-typeable primers specific to the amplification of this biallelic marker or group of biallelic markers. The mutation can be identified by conducting sequencing reactions on the nucleic acids which hybridize with the probes defined herein or which show amplification by PCR.

The region of the *TBC-1* gene containing the mutation responsible for the detectable phenotype may be used in diagnostic techniques such as those described below. For example, microsequencing oligonucleotides, or oligonucleotides containing the mutation responsible for the detectable phenotype for amplification, or hybridization based diagnostics, such as those described herein, may be used for detecting individuals suffering from the detectable phenotype or individuals at risk of developing the detectable phenotype at a subsequent time. In addition, the *TBC-1* allele responsible for the detectable phenotype may be used in gene therapy. The *TBC-1* allele responsible for the detectable phenotype may also be cloned into an expression vector to express the mutant TBC-1 protein as described herein.

F- DETECTION OF MARKERS OR GROUPS OF MARKERS ASSOCIATED WITH A TRAIT

The invention concerns a diagnostic method for the detection in an individual of markers or groups of markers associated with a trait which may include prostate cancer susceptibility, an early onset of prostate cancer, or the aggressiveness of prostate cancer tumors. The information obtained using this method is useful in the diagnosis, staging, monitoring, prognosis and/or prophylactic or curative therapy of prostate cancer. The method also concerns the detection of specific alleles present within a gene expressing a modified level of *TBC-1* mRNA or an altered *TBC-1* mRNA, coding for an altered TBC-1 protein. More particularly, the invention concerns the detection of a *TBC-1* nucleic acid comprising at least one of the nucleotide sequences of SEQ ID Nos 7-8.

This method comprises the following steps :

- obtaining a nucleic acid sample from the individual to be tested, and
- determining the presence in the sample of an allele of a biallelic marker or of a group of biallelic markers of the *TBC-1* gene which, when taken alone or in combination with another/other biallelic marker/s of the *TBC-1* gene, is indicative of prostate cancer, of an early

onset of prostate cancer, of the level of aggressiveness of prostate cancer tumors, of a modified or forthcoming expression of the *TBC-1* gene, of a modified or forthcoming production of the TBC-1 protein, or of the production of a modified TBC-1 protein.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8, or a fragment thereof including the polymorphic base.

More particularly, the detection method of the present invention comprises the following steps:

obtaining a nucleic acid sample from the individual to be tested,

- amplifying a nucleotide sequence of the *TBC-1* gene contained in the sample, and
- detecting the presence in the sample of an allele of a biallelic marker or of a group of biallelic markers of the *TBC-1* gene which, when taken alone or in combination with another/other biallelic marker/s of the *TBC-1* gene, is indicative of prostate cancer of an early onset of prostate cancer, of the level of aggressiveness of prostate cancer tumors, of a modified or forthcoming expression of the *TBC-1* gene, of a modified or forthcoming production of the TBC-1 protein, or of the production of a modified TBC-1 protein.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8. In preferred embodiments of the two detection methods described above, the presence of alleles of one or more biallelic markers of the *TBC-1* gene is determined through microsequencing reactions using microsequencing primers such as those of SEQ ID Nos 11-12. More particularly, it is preferred that the microsequencing primers be bound to a solid support, preferably in the form of arrays of primers attached to appropriate chips or be used in microfluidic devices. Such arrays are described in further detail in the next section.

Alternatively, the detection method of the present invention can comprise the following steps:

- obtaining a nucleic acid sample from the individual to be tested,
- specifically amplifying a nucleotide sequence of the *TBC-1* gene comprising an allele of biallelic marker or of a group of biallelic markers of a *TBC-1* gene which, when taken alone or in combination with another/other biallelic marker/s of a *TBC-1* gene, is indicative of prostate cancer, of an early onset of prostate cancer, of the level of aggressiveness of prostate cancer tumors, of a modified or forthcoming expression of the *TBC-1* gene, of a modified or forthcoming production of the TBC-1 protein, or of the production of a modified TBC-1 protein, and
- detecting the presence of an amplification product.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

In a further embodiment of the present invention, another detection method comprises the following steps:

- contacting a hybridization probe in accordance with the invention with a biological sample, the DNA of which can optionally be pretreated to be rendered accessible to hybridization, under conditions which allow hybridization of the probe to the nucleotide sequence of the *TBC-1* gene comprising an allele of a biallelic marker of the present invention, and
- revealing the formation of a hybrid comprising the probe and a nucleotide sequence from the biological sample.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

Preferably, the formation of a hybrid is revealed through the detection of a signal from a label attached to the probe. More preferably, the signal is amplified prior to being revealed.

In yet another embodiment of the present invention, a further detection method comprises the following steps:

- contacting a hybridization probe in accordance with the invention, immobilized on a solid support, with a biological sample, the DNA of which can optionally be pretreated to be rendered accessible to hybridization, under conditions which allow hybridization of the probe to the nucleotide sequence of the *TBC-1* gene comprising an allele of a biallelic marker of the present invention,
- contacting the hybrid thus formed, optionally after having removed the nucleotide sequences which did not hybridize with the probe, with a labeled hybridization probe, and
- revealing the formation of a hybrid comprising the hybridization probe, a nucleotide sequence from the biological sample and the labeled probe.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

Preferably, the nucleotide sequences of the biological sample are amplified prior to hybridization using one of the primers described herewith.

The invention also specifically relates to a method of determining whether an individual suffering from prostate cancer or susceptible of developing prostate cancer is likely to respond positively to treatment with a selected medicament acting against prostate cancer.

The method comprises the following steps:

- obtaining a DNA sample from the individual to be tested, and
- analyzing said DNA sample to determine whether it comprises alleles of one or more biallelic markers associated with a positive response to treatment with the medicament and/or alleles of one or more biallelic markers associated with a negative response to treatment with the medicament.

The detection methods of the present invention can be applied to, for example, the preliminary screening of patient populations suffering from prostate cancer. This preliminary screening is useful to initiate adequate treatment when needed or to determine and select appropriate patient populations for clinical trials on new compounds in order to avoid the potential occurrence of specific side effects or to enhance the probability of beneficial patient response. By establishing in advance a homogeneous genotype selection for the population to be tested, the assessment of drug efficacy and/or toxicity can be more readily achieved and less hampered by divergences in population response. This approach can yield better therapeutic approaches based on patient population targeting resulting from pharmacogenomics studies.

Further details on the two main process steps of this method are provided below.

Nucleic acid samples

A nucleic acid sample is obtained from an individual to be tested for susceptibility to one of the traits referred to above. The nucleic acid sample may be an RNA sample or a DNA sample, but is preferably genomic DNA. Any human-derived sample containing the nucleic acid of interest can be employed without any particular limitation. As examples, there may be mentioned body fluids such as blood, bone marrow fluid, semen and peritoneal fluid, prostate and other tissue cells.

Appropriate genomic DNA can be prepared by extracting and purifying from these samples by conventional methods. Concentrated test samples can also be obtained in large quantities by amplifying, based on the genomic DNA, a region susceptible of containing a biallelic marker or a group of biallelic markers as described above, or biallelic markers in LD therewith, responsible for the given trait. This is carried out, for example by the PCR technique, employing primers adequately designed so that the desired region is amplified.

Preferably, the nucleic acid sample subject to analysis in this method is genomic and in such a case, an amplifying step may be applied to the nucleic acid sample, primed from a forward primer sufficiently complementary with a first subsequence of the antisense strand of the *TBC-1* gene to hybridize therewith, and a reverse primer sufficiently complementary to a second sequence of the sense strand of the *TBC-1* gene to hybridize therewith. Amplification products generated from these primers are analyzed using methods known to the skilled technician and which are described in further detail in the next section.

More particularly, it is desirable to amplify genomic DNA regions susceptible of harboring a biallelic marker such as that disclosed in example 4.

The preferred primers that can be used to amplify such regions are disclosed in example 3.

Detection of biallelic markers in the amplified nucleic acid samples

Once amplification of the DNA has been carried out or once the appropriate DNA sample has been obtained if no amplification takes place, the identities of the polymorphic bases of one or more of the biallelic markers described above or biallelic markers in with the sequences of such markers, or a combination thereof, are determined. The identities of the polymorphic bases may be determined using either any of the genotyping procedures described above in "Genotyping of biallelic markers".

For example, microsequencing with any primer having a 3' end near the polymorphic nucleotide, and preferably immediately adjacent to the polymorphic nucleotide, may be used. Alternatively, the PCR product may be completely sequenced to determine the identities of the polymorphic bases in the biallelic markers. In another method, the identities of the polymorphic bases in the biallelic markers is determined by hybridizing the amplification products to allele specific oligonucleotides specific for the polymorphic bases in the biallelic markers, such as the probes described previously.

More preferably, the identities of the polymorphic bases can be determined with "large scale" genotyping procedures, such as those involving the hybridization with DNA chips or integrated microsequencing systems, which are described above. An example of the microsequencing analysis alternative on oligonucleotide arrays is described in Pastinen et al. (1997), the disclosure of which is incorporated herein by reference. In another example, microarrays (DNA chips) comprising allele specific oligonucleotides can be used and are described above.

It will be appreciated that the identities of the polymorphic bases in the biallelic markers may be determined using techniques other than those listed above, such as conventional dot blot analyses.

The invention also relates to diagnostic kits useful for determining the presence in a DNA sample of alleles associated with a prostate cancer susceptibility, or the aggressiveness of prostate cancer tumors, or more generally of a modified TBC-1 production and/or metabolism, preferably of a production of an altered TBC-1 protein.

In a first embodiment, the kit comprises primers such as those described above, preferably forward and reverse primers which are used to amplify the *TBC-1* gene, particularly the genomic sequence of SEQ ID No1, or a fragment thereof. In some embodiments, at least one of the primers is complementary to a nucleotide sequence of the *TBC-1* gene comprising a biallelic marker associated with prostate cancer, with an early onset of prostate cancer, with the level of aggressiveness of prostate cancer tumors, with a modified or forthcoming expression of the *TBC-1* gene, with a modified or forthcoming production of the TBC-1 protein, or with the production of a modified TBC-1 protein. In one embodiment, the biallelic marker comprises one of the sequences of SEQ ID Nos 7-8. In a preferred embodiment, the kit comprises one or more of the sequences of SEQ ID Nos 9-10.

In a second embodiment, the kit comprises microsequencing primers, preferably those of SEQ ID Nos 11-12.

In a third embodiment, the kit comprises primers for the amplification of the first or second *TBC-1* cDNAs, such as the sequences of SEQ ID No 13 and 14.

In a fourth embodiment, the kit comprises a primer which is complementary to any nucleotide sequence of the *TBC-1* gene, particularly of the genomic sequence of SEQ ID No 1, and is used to amplify the *TBC-1* gene or a fragment thereof contained in the nucleic acid sample to be tested which includes a polymorphic base of at least one biallelic marker. Preferably, the amplified region includes a polymorphic base of at least one biallelic marker selected from the group consisting of SEQ ID Nos 7-8. In some embodiments, the primer comprises one of the sequences of SEQ ID Nos 9-10.

In a fifth embodiment, the kit comprises a DNA probe, that is or eventually becomes immobilized on a solid support, which is capable of hybridizing with the *TBC-1* gene, preferably with a region of the *TBC-1* gene which comprises a biallelic marker of the present invention. The techniques for immobilizing a nucleotide primer or probe on a solid support are well-known to the skilled person and include, but are not limited to, the immobilization techniques described in the present application.

The kits of the present invention can also comprise optional elements including appropriate amplification reagents such as DNA polymerases when the kit comprises primers, reagents useful in hybridization reactions and reagents useful to reveal the presence of a hybridization reaction between a labeled hybridization probe and the *TBC-1* gene containing at least one biallelic marker. In one embodiment, the biallelic marker comprises one of the sequences of SEQ ID Nos 7-8.

G- TREATMENT OF PROSTATE CANCER

As the metastasis of prostate cancer can be fatal, it is important to detect prostate cancer susceptibility of individuals. Consequently, the invention also concerns a method for the treatment of prostate cancer comprising the following steps:

- selecting an individual whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with prostate cancer;
- following up said individual for the appearance (and optionally the development) of tumors in prostate; and
- administering a treatment acting against prostate cancer to said individual at an appropriate stage of the prostate cancer.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

Another embodiment of the present invention consists of a method for the treatment of prostate cancer comprising the following steps:

- selecting an individual whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with prostate cancer;
- administering a preventive treatment of prostate cancer such as 4HPR to said individual.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

In a further embodiment, the present invention concerns a method for the treatment of prostate cancer comprising the following steps:

- selecting an individual whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with prostate cancer;
- administering a preventive treatment of prostate cancer such as 4HPR to said individual;
- following up said individual for the appearance and the development of tumors in prostate; and optionally
- administering a treatment acting against prostate cancer to said individual at the appropriate stage of the prostate cancer.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

To enlighten the choice of the appropriate beginning of the treatment of prostate cancer, the present invention also concerns a method for the treatment of prostate cancer comprising the following steps:

- selecting an individual suffering from a prostate cancer whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with the aggressiveness of prostate cancer tumors; and
- administering a treatment acting against prostate cancer to said individual.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

The invention also concerns a method for the treatment of prostate cancer in a selected population of individuals. The method comprises :

- selecting an individual suffering from prostate cancer and whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with a positive response to treatment with an effective amount of a medicament acting against prostate cancer,
- and/or whose DNA does not comprise alleles of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with a negative response to treatment with said medicament; and
- administering at suitable intervals an effective amount of said medicament to said selected individual.

In some embodiments, the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

In the context of the present invention, a "positive response" to a medicament can be defined as comprising a reduction of the symptoms related to the disease, an increase of survival time or condition to be treated.

In the context of the present invention, a "negative response" to a medicament can be defined as comprising either a lack of positive response to the medicament which does not lead to a symptom reduction or an increase of survival time, or which leads to a side-effect observed following administration of the medicament.

The invention also relates to a method of determining whether a subject is likely to respond positively to treatment with a medicament.

The method comprises identifying a first population of individuals who respond positively to said medicament and a second population of individuals who respond negatively to said medicament. One or more biallelic markers is identified in the first population which is associated with a positive response to said medicament or one or more biallelic markers is identified in the second population which is associated with a negative response to said medicament. The biallelic markers may be identified using the techniques described herein.

A DNA sample is then obtained from the subject to be tested. The DNA sample is analyzed to determine whether it comprises alleles of one or more biallelic markers associated with a positive response to treatment with the medicament and/or alleles of one or more biallelic markers associated with a negative response to treatment with the medicament.

In some embodiments, the medicament may be administered to the subject in a clinical trial if the DNA sample contains alleles of one or more biallelic markers associated with a positive response to treatment with the medicament and/or if the DNA sample lacks alleles of one or more biallelic markers associated with a negative response to treatment with the medicament. In preferred embodiments, the medicament is a drug acting against prostate cancer. In other embodiments, the biallelic marker has a sequence selected from the group consisting of SEQ ID Nos 7-8.

Using the method of the present invention, the evaluation of drug efficacy may be conducted in a population of individuals likely to respond favorably to the medicament.

Another aspect of the invention is a method of using a medicament comprising obtaining a DNA sample from a subject, determining whether the DNA sample contains alleles of one or more biallelic markers associated with a positive response to the medicament and/or whether the DNA sample contains alleles of one or more biallelic markers associated with a negative response to the medicament, and administering the medicament to the subject if the DNA sample contains alleles of one or more biallelic markers associated with a positive response to the medicament and/or if the DNA sample lacks alleles of one or more biallelic markers associated with a negative response to the medicament.

The invention also concerns a method for the clinical testing of a medicament, preferably a medicament acting against prostate cancer. The method comprises the following steps:

- administering a medicament, preferably a medicament susceptible of acting against prostate cancer to a heterogeneous population of individuals,
- identifying a first population of individuals who respond positiveiy to said medicament and a second population of individuals who respond negatively to said medicament,

identifying biallelic markers in said first population which are associated with a positive response to said medicament,

- selecting individuals whose DNA comprises biallelic markers associated with a positive response to said medicament, and
- administering said medicament to said individuals.

Such methods are deemed to be extremely useful to increase the benefit/risk ratio resulting from the administration of medicaments which may cause undesirable side effects and/or be inefficacious to a portion of the patient population to which it is normally administered.

Once an individual has been diagnosed as suffering from a prostate cancer, selection tests are carried out to determine whether the DNA of this individual comprises alleles of a biallelic marker or of a group of biallelic markers associated with a positive response to treatment or with a negative response to treatment which may include either side effects or unresponsiveness.

The selection of the patient to be treated using the method of the present invention can be carried out through the detection methods described above. The individuals which are to be selected are preferably those whose DNA does not comprise alleles of a biallelic marker or of a group of biallelic markers associated with a negative response to treatment. The knowledge of

an individual's genetic predisposition to unresponsiveness or side effects to particular medicaments allows the clinician to direct treatment toward appropriate drugs against prostate cancer.

Once the patient's genetic predispositions have been determined, the clinician can select appropriate treatment for which negative response, particularly side effects, has not been reported or has been reported only marginally for the patient.

H. EXPRESSION OF A TBC-1 REGULATORY OR CODING POLYNUCLEOTIDE OF THE INVENTION.

Any of the regulatory polynucleotides or the coding polynucleotides of the invention may be inserted into recombinant vectors for expression in a recombinant host cell or a recombinant host organism.

Thus, the present invention also encompasses a family of recombinant vectors that contains either a regulatory polynucleotide selected from the group consisting of any one of the regulatory polynucleotides derived from the *TBC-1* genomic sequence of SEQ ID No 1, or a polynucleotide comprising the *TBC-1* coding sequence, or both.

In a first preferred embodiment, a recombinant vector of the invention is used as an expression vector : (a) the *TBC-1* regulatory sequence comprised therein drives the expression of a coding polynucleotide operably linked thereto; (b) the *TBC-1* coding sequence is operably linked to regulation sequences allowing its expression in a suitable cell host and/or host organism.

In a second preferred embodiment, a recombinant vector of the invention is used to amplify the inserted polynucleotide derived from the *TBC-1* genomic sequence of SEQ ID No 1 or *TBC-1* cDNAs in a suitable cell host , this polynucleotide being amplified at every time that the recombinant vector replicates.

More particularly, the present invention relates to expression vectors which include nucleic acids encoding a TBC-1 protein, preferably the TBC-1 protein of the amino acid sequence of SEQ ID No 5 described therein, under the control of a regulatory sequence selected among the *TBC-1* regulatory polynucleotides, or alternatively under the control of an exogenous regulatory sequence.

A recombinant expression vector comprising a nucleic acid selected from the group consisting of SEQ ID No 2, or biologically active fragments or variants thereof, is also part of the present invention.

The invention also encompasses a recombinant expression vector comprising :

a) a nucleic acid comprising a regulatory polynucleotide of the nucleotide sequence SEQ ID No 2, or a biologically active fragment or variant thereof;

b) a polynucleotide encoding a polypeptide or a polynucleotide of interest operably linked with said nucleic acid.

c) optionally, a nucleic acid comprising a 3'-regulatory polynucleotide, preferably a 3'-regulatory polynucleotide of the invention, or a biologically active fragment or variant thereof.

The nucleic acid comprising the nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof may also comprises the 5'-UTR sequence located between the nucleotide at position 1 and the nucleotide at position 170 of SEQ ID No 3, or a biologically active fragment or variant thereof.

The nucleic acid comprising the nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof may also comprises the 5'-UTR sequence located between the nucleotide at position 1 and the nucleotide at position 175 of SEQ ID No 4, or a biologically active fragment or variant thereof.

The invention also pertains to a recombinant expression vector useful for the expression of the *TBC-1* coding sequence, wherein said vector comprises a nucleic acid selected from the group consisting of SEQ ID Nos 3 and 4 or a nucleic acid having at least 95% nucleotide identity with a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4.

Another recombinant expression vector of the invention consists in a recombinant vector comprising a nucleic acid comprising the nucleotide sequence beginning at the nucleotide in position 176 and ending in position 3733 of the polynucleotide of SEQ ID No 4.

Some of the elements which can be found in the vectors of the present invention are described in further detail in the following sections.

a) Vectors

A recombinant vector according to the invention comprises, but is not limited to, a YAC (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a cosmid, a plasmid or even a linear DNA molecule which may consist of a chromosomal, non-chromosomal and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of :

(1) a genetic element or elements having a regulatory role in gene expression, for example promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp in length that act on the promoter to increase the transcription.

(2) a structural or coding sequence which is transcribed into mRNA and eventually translated into a polypeptide, and

(3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where a recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

Generally, recombinant expression vectors will include origins of replication, selectable markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably a leader sequence capable of directing secretion of the translated protein into the periplasmic space or the extracellular medium.

The selectable marker genes for selection of transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicin or ampicillin resistance in *E. coli*, or levan saccharase for mycobacteria.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and a bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial

vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and GEM1 (Promega Biotec, Madison, WI, USA).

Large numbers of suitable vectors and promoters are known to those of skill in the art, and commercially available, such as bacterial vectors : pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); or eukaryotic vectors : pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); baculovirus transfer vector pVL1392/1393 (Pharmingen); pQE-30 (QIAexpress).

A suitable vector for the expression of the TBC-1 polypeptide of SEQ ID No 5 is a baculovirus vector that can be propagated in insect cells and in insect cell lines. A specific suitable host vector system is the pVL1392/1393 baculovirus transfer vector (Pharmingen) that is used to transfect the SF9 cell line (ATCC N°CRL 1711) which is derived from *Spodoptera frugiperda*.

Other suitable vectors for the expression of the TBC-1 polypeptide of SEQ ID No 5 in a baculovirus expression system include those described by Chai et al. (1993), Vlasak et al. (1983) and Lenhard et al. (1996).

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example SV40 origin, early promoter, enhancer, splice and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

b) Promoters

The suitable promoter regions used in the expression vectors according to the present invention are chosen taking into account the cell host in which the heterologous gene has to be expressed.

A suitable promoter may be heterologous with respect to the nucleic acid for which it controls the expression or alternatively can be endogenous to the native polynucleotide containing the coding sequence to be expressed. Additionally, the promoter is generally heterologous with

respect to the recombinant vector sequences within which the construct promoter/coding sequence has been inserted.

Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen) (Smith et al., 1983; O'Reilly et al., 1992), the lambda P_R promoter or also the trc promoter.

Promoter regions can be selected from any desired gene using, for example, CAT (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors. Particularly preferred bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, PL and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art.

The choice of a promoter is well within the ability of a person skilled in the field of genetic egineering. For example, one may refer to the book of Sambrook et al. (1989) or also to the procedures described by Fuller et al. (1996).

The vector containing the appropriate DNA sequence as described above, more preferably a *TBC-1* gene regulatory polynucleotide, a polynucleotide encoding the TBC-1 polypeptide of SEQ ID No 5 or both of them, can be utilized to transform an appropriate host to allow the expression of the desired polypeptide or polynucleotide.

c) Other types of vectors

The *in vivo* expression of a TBC-1 polypeptide of SEQ ID No 5 may be useful in order to correct a genetic defect related to the expression of the native gene in a host organism or to the production of a biologically inactive TBC-1 protein.

Consequently, the present invention also deals with recombinant expression vectors mainly designed for the *in vivo* production of the TBC-1 polypeptide of SEQ ID No 5 by the introduction of the appropriate genetic material in the organism of the patient to be treated. This genetic material may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell being subsequently reintroduced in the said organism, directly *in vivo* into the appropriate tissue.

By « vector » according to this specific embodiment of the invention is intended either a circular or a linear DNA molecule.

One specific embodiment for a method for delivering a protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide of interest into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect.

In a specific embodiment, the invention provides a composition for the *in vivo* production of the TBC-1 protein or polypeptide described herein. It comprises a naked polynucleotide operatively coding for this polypeptide, in solution in a physiologically acceptable carrier, and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

Compositions comprising a polynucleotide are described in PCT application N° WO 90/11092 (Vical Inc.) and also in PCT application N° WO 95/11307 (Institut Pasteur, INSERM, Université d'Ottawa) as well as in the articles of Tacson et al. (1996) and of Huygen et al. (1996).

The amount of vector to be injected to the desired host organism varies according to the site of injection. As an indicative dose, it will be injected between 0,1 and 100 µg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, it may be introduced in vitro in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired TBC-1 polypeptide or the desired fragment thereof is reintroduced into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

In one specific embodiment, the vector is derived from an adenovirus. Preferred adenovirus vectors according to the invention are those described by Feldman and Steg (1996) or Ohno et al. (1994). Another preferred recombinant adenovirus according to this specific embodiment of the present invention is the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin (French patent application N° FR-93.05954).

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery systems of choice for the transfer of exogenous polynucleotides *in vivo* , particularly to mammals, including humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host

Particularly preferred retroviruses for the preparation or construction of retroviral *in vitro* or *in vitro* gene delivery vehicles of the present invention include retroviruses selected from the group consisting of Mink-Cell Focus Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis virus and Rous Sarcoma virus. Particularly preferred Murine Leukemia Viruses include the 4070A and the 1504A viruses, Abelson (ATCC No VR-999), Friend (ATCC No VR-245), Gross (ATCC No VR-590), Rauscher (ATCC No VR-998) and Moloney Murine Leukemia Virus (ATCC No VR-190; PCT Application No WO 94/24298). Particularly preferred Rous Sarcoma Viruses include Bryan high titer (ATCC Nos VR-334, VR-657, VR-726, VR-659 and VR-728). Other preferred retroviral vectors are those described in Roth et al. (Roth J.A. et al., 1996), PCT Application No WO 93/25234, PCT Application No WO 94/ 06920, Roux et al., 1989, Julan et al., 1992 and Neda et al., 1991.

Yet another viral vector system that is contemplated by the invention consists in the adeno-associated virus (AAV). The adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka et al., 1992). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (Flotte et al., 1992; Samulski et al., 1989; McLaughlin et al., 1989). One advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

Other compositions containing a vector of the invention advantageously comprise an oligonucleotide fragment of a nucleic sequence selected from the group consisting of SEQ ID Nos 2 or 3 as an antisense tool that inhibits the expression of the corresponding *TBC-1* gene. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al. (1995) or those described in PCT Application No WO 95/24223.

Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to the 5'end of the *TBC-1* mRNAs. In another embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of *TBC-1* that contains the translation initiation codon ATG.

Host cells

Another object of the invention consists in host cell that have been transformed or transfected with one of the polynucleotides described therein, and more precisely a polynucleotide either comprising a *TBC-1* regulatory polynucleotide or the coding sequence of the TBC-1 polypeptide having the amino acid sequence of SEQ ID No 5. Are included host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above.

A recombinant host cell of the invention comprises any one of the polynucleotides or the recombinant vectors described therein.

A preferred recombinant host cell according to the invention comrprises a polynucleotide selected from the following group of polynucleotides :

a) a purified or isolated nucleic acid encoding a TBC-1 polypeptide, or a polypeptide fragment or variant thereof.

b) a purified or isolated nucleic acid comprising at least 20 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4.

c) a purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID No 2 or a biologically active fragment or variant of the nucleotide sequence of SEQ ID No 2.

d) a purified or isolated nucleic acid comprising a 3'-regulatory sequence of the *TBC-1* gene, or a biologically active fragment or variant thereof.

e) a polynucleotide consisting of :

(1) a nucleic acid comprising a regulatory polynucleotide of SEQ ID No 2 or a biologically active fragment or variant thereof;

(2) a polynucleotide encoding a desired polypeptide or nucleic acid.

(3) Optionally, a nucleic acid comprising a 3'-regulatory sequence , preferably a 3'-regulatory sequence of the *TBC-1* gene, or a biologically active fragment or variant thereof, wherein sequences (1), (2) and (3) are operably linked to one other.

Another preferred recombinant cell host according to the present invention is characterized in that its genome or genetic background (including chromosome, plasmids) is modified by the nucleic acid coding for the TBC-1 polypeptide of SEQ ID No 5.

Preferred host cells used as recipients for the expression vectors of the invention are the following :

a) Prokaryotic host cells : *Escherichia coli* strains (I.E. DH5-α strain) or *Bacillus subtilis*.

b) Eukaryotic host cells : HeLa cells (ATCC N°CCL2; N°CCL2.1; N°CCL2.2), Cv 1 cells (ATCC N°CCL70), COS cells (ATCC N°CRL1650; N°CRL1651), Sf-9 cells (ATCC N°CRL1711).

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in the expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skill artisan.

Transgenic animals

The terms "transgenic animals" or "host animals" are used herein to designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising a *TBC-1* coding sequence, a *TBC-1* regulatory polynucleotide or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

Preferred transgenic animals according to the invention contain in their somatic cells and/or in their germ line cells a polynucleotide selected from the following group of polynucleotides :

a) a purified or isolated nucleic acid encoding a TBC-1 polypeptide, or a polypeptide fragment or variant thereof.

b) a purified or isolated nucleic comprising at least 20 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4.

c) a purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID No 2 or a biologically active fragment or variant of the nucleotide sequence of SEQ ID No 2.

d) a purified or isolated nucleic acid comprising a 3'-regulatory sequence of the *TBC-1* gene, or a biologically active fragment or variant thereof.

e) a polynucleotide consisting of :

(1) a nucleic acid comprising a regulatory polynucleotide of SEQ ID No 2 or a biologically active fragment or variant thereof;

(2) a polynucleotide encoding a desired polypeptide or nucleic acid.

(3) Optionally, a nucleic acid comprising a 3'-regulatory sequence, preferably a 3'regulatory sequence of the *TBC-1* gene, or a biologically active fragment or variant thereof, wherein sequences (1), (2) and (3) are operably linked to one other.

The transgenic animals of the invention thus contain specific sequences of exogenous genetic material such as the nucleotide sequences described above in detail.

In a first preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to cell differentiation, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native TBC-1 protein, or alternatively a mutant TBC-1 protein.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of the regulatory polynucleotides of the *TBC-1* gene, leading to good

yields in the synthesis of this protein of interest, and eventually a tissue specific expression of this protein of interest.

Since it is possible to produce transgenic animals of the invention using a variety of different sequences, a general description will be given of the production of transgenic animals by referring generally to exogenous genetic material. This general description can be adapted by those skilled in the art in order to incorporate the DNA sequences into animals. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to Sandou et al. (1994) and also to US Patents Nos 4,873,191, issued Oct.10, 1989, 5,968,766, issued Dec. 16, 1997 and 5,387,742, issued Feb. 28, 1995, these documents being herein incorporated by reference to disclose methods for producing transgenic mice.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that incorporates exogenous genetic material which is integrated into the genome. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a *TBC-1* coding sequence, a *TBC-1* regulatory polynucleotide or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is made using electroporation. The cells subjected to electroporation are screened (e.g. Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour et al. (1988). Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice. The blastocysts are then inserted into a female host animal and allowed to grow to term. The offsprings of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

I. TBC-1 POLYPEPTIDE AND PEPTIDE FRAGMENTS

It is now easy to produce proteins in high amounts by genetic engineering techniques through expression vectors such as plasmids, phages or phagemids. The polynucleotide that code for one

the polypeptides of the present invention is inserted in an appropriate expression vector in order to produce the polypeptide of interest *in vitro*.

Thus, the present invention also concerns a method for producing one of the polypeptides described herein, and especially a polypeptide of SEQ ID No 5 or a fragment or a variant thereof, wherein said method comprises the steps of :

a) culturing, in an appropriate culture medium, a cell host previously transformed or transfected with the recombinant vector comprising a nucleic acid encoding a TBC- polypeptide, or a fragment or a variant thereof;

b) harvesting the culture medium thus conditioned or lyse the cell host, for example by sonication or by an osmotic shock;

c) separating or purifying, from the said culture medium, or from the pellet of the resultant host cell lysate the thus produced polypeptide of interest.

d) Optionally characterizing the produced polypeptide of interest.

In a specific embodiment of the above method, step a) is preceded by a step wherein the nucleic acid coding for a TBC-1 polypeptide, or a fragment or a variant thereof, is inserted in an appropriate vector, optionally after an appropriate cleavage of this amplified nucleic acid with one or several restriction endonucleases. The nucleic acid coding for a TBC-1 polypeptide or a fragment or a variant thereof may be the resulting product of an amplification reaction using a pair of primers according to the invention (by SDA, TAS, 3SR NASBA, TMA etc.).

The polypeptides according to the invention may be characterized by binding onto an immunoaffinity chromatography column on which polyclonal or monoclonal antibodies directed to a polypeptide of SEQ ID No 5, or a fragment or a variant thereof, have previously been immobilized.

Purification of the recombinant proteins or peptides according to the present invention may be carried out by passage onto a Nickel or Cupper affinity chromatography column. The Nickel chromatography column may contain the Ni-NTA resin (Porath et al., 1975).

The polypeptides or peptides thus obtained may be purified, for example by high performance liquid chromatography, such as reverse phase and/or cationic exchange HPLC, as described by Rougeot et al. (1994). The reason to prefer this kind of peptide or protein purification is the lack of byproducts found in the elution samples which renders the resultant purified protein or peptide more suitable for a therapeutic use.

Another object of the present invention consists in a purified or isolated TBC-1 polypeptide or a fragment or a variant thereof.

In a preferred embodiment, the TBC-1 polypeptide comprises an amino acid sequence of SEQ ID No 5 or a fragment or a variant thereof.

The TBC-1 polypeptide of the invention possesses amino acid homologies as regards to the murine TBC-1 protein of 1141 amino acids in length which is described in US Patent No US 5,700,927. The TBC-1 protein of the invention also possesses some homologies with two other proteins : the Pollux drosophila protein (Zhang et al., 1996) and the CDC16 protein from *Caenorhabditis elegans* (Wilson et al., 1994). Figure 2 represents an amino acid alignment of a portion of the amino acid sequence of the TBC-1 protein of SEQ ID No 5 with other proteins sharing amino acid homology with TBC-1. The upper line shows the whole amino acid sequence of the murine tbc-1 protein described in US Patent No US 5,700,927; the second line represents part of the amino acid sequence of the TBC-1 protein of SEQ ID No 5; the third line (Genbank access No : *dmu50542*) depicts the amino acid sequence of the Pollux protein mentioned above; the fourth line (Genbank access No : *celf35h12*) shows the amino acid sequence of the *C. elegans* protein mentioned above; the fifth line presents positions in which consensus amino acids are identified, i.e. amino acids shared by the sequences presented in the four upper lines, when present.

The TBC-1 polypeptide of the amino acid sequence of SEQ ID No 5 has 1185 amino acids in length. The TBC-1 polypeptide includes a "TBC domain" which is spanning from the amino acid in position 786 to the amino acid in position 974 of the amino acid sequence of SEQ ID No 5. This TBC domain is represented in Figure 2 as a grey area spanning from the amino acid numbered 758 to the amino acid numbered 949. This TBC domain is likely to regulate protein-protein interactions. Moreover, the TBC-1 TBC domain includes the amino acid sequence EVGYCQGL, spanning from the amino acid in position 886 to the amino acid in position 893 of the amino acid sequence of SEQ ID No 5. The EVGYCQGL amino acid sequence spans from the amino acid numbered 861 to the amino acid numbered 868 of Figure 2. This site may interact with a kinase. Based on the structural similarity to *cdc16*, a yeast regulator of mitosis, TBC-1 is likely to regulate mitosis and cytokinesis by interacting with other proteins which also participate with the regulation of mitosis, cytokinesis and septum formation.

Preferred polypeptides of the invention comprise the TBC domain of TBC-1, or alternatively at least the EVGYCQGL amino acid sequence motif.

A further object of the present invention concerns a purified or isolated polypeptide which is encoded by a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 1, 3, and 4 or fragments or variants thereof.

In a preferred embodiment, a variant TBC-1 polypeptide comprises amino acid changes ranging from 1, 2, 3 , 4, 5, 10 to 20 substitutions, additions or deletions of one aminoacid, preferably from 1 to 10, more preferably from 1 to 5 and most preferably from 1 to 3 substitutions, additions or deletions of one amino acid. The preferred amino acid changes are those which have little or no influence on the biological activity or the capacity of the variant TBC-1 polypeptide to be recognized by antibodies raised against a native TBC-1 protein.

A single variant molecule of the TBC-1 protein is explicitly excluded from the scope of the present invention, which is a polypeptide having the same amino acid sequence than the murine tbc1 protein described in the US Patent No 5,700,927.

In a second preferred embodiment, a mutated TBC-1 polypeptide comprises amino acid changes ranging from 1 to about 200 deletions of one amino acid and of at least one aminoacid substitution or addition, preferably from 1 to 10, 20 or 30 amino acid substitutions or additions. The amino acid substitutions are generally non conservative in terms of polarity, charge, hydrophilicity properties of the substitute amino acid when compared with the native amino acid. The amino acid changes occurring in such a mutated TBC-1 polypeptide may be determinant for the biological activity or for the capacity of the mutated TBC-1 polypeptide to be recognized by antibodies raised against a native TBC-1.

Amino acid deletions, additions or substitutions in the TBC-1 protein are preferably located outside of the TBC domain as defined above. Most preferably, a mutated TBC-1 protein has an intact "EVGYCQGL" amino acid motif.

Such a mutated TBC-1 protein may be the target of diagnostic tools, such as specific monoclonal or polyclonal antibodies, useful for detecting the mutated TBC-1 protein in a sample.

In the case of an aminoacid substitution in the amino acid sequence of a polypeptide acording to the invention, one or several -consecutive or non-consecutive- aminoacids are replaced by « equivalent » aminoacids. The expression « equivalent » aminoacid is used herein to designate any aminoacid that may be substituted for one of the aminoacids belonging to the native protein structure without decreasing the binding properties of the corresponding peptides to the antibodies raised against the TBC-1 protein of the amino acid sequence of SEQ ID No 5. In other words, the «equivalent » aminoacids are those which allow the generation or the synthesis of a polypeptide with a modified sequence when compared to the aminoacid sequence of the native TBC-1 protein, said modified polypeptide being able to bind to the antibodies raised against the TBC-1 protein of the amino acid sequence of SEQ ID No 5 and/or to induce antibodies recognizing the parent polypeptide consisting in the TBC-1 polypeptide of the amino acid sequence of SEQ ID No 5.

These equivalent aminoacids may be determined either by their structural homology with the initial aminoacids to be replaced, by the similarity of their net charge, and optionally by the results of the cross-immunogenicity between the parent peptides and their modified counterparts.

The peptides containing one or several « equivalent » aminoacids must retain their specificity and affinity properties to the biological targets of the parent protein, as it can be assessed by a ligand binding assay or an ELISA assay.

By an equivalent aminoacid according to the present invention is also meant the replacement of a residue in the L-form by a residue in the D form or the replacement of a Glutamic acid (E) residue by a Pyro-glutamic acid compound. The synthesis of peptides containing at least one residue in the D-form is, for example, described by Koch (1977).

A specific embodiment of a modified TBC-1 peptide molecule of interest according to the present invention, includes, but is not limited to, a peptide molecule which is resistant to proteolysis, is a peptide in which the -CONH- peptide bond is modified and replaced by a (CH_2NH) reduced bond, a (NHCO) retro inverso bond, a (CH_2-O) methylene-oxy bond, a (CH_2-S) thiomethylene bond, a (CH_2CH_2) carba bond, a (CO-CH_2) cetomethylene bond, a (CHOH-CH_2) hydroxyethylene bond), a (N-N) bound, a E-alcene bond or also a -CH=CH- bond.

The invention also encompasses a TBC-1 polypeptide or a fragment or a variant thereof in which at least one peptide bound has been modified as described above.

The polypeptides according to the invention may also be prepared by the conventional methods of chemical synthesis, either in a homogenous solution or in solid phase. As an illustrative embodiment of such chemical polypeptide synthesis techniques, it may be cited the homogenous solution technique described by Houbenweyl (1974).

The TBC-1 polypeptide, or a fragment or a variant thereof may thus be prepared by chemical synthesis in liquid or solid phase by successive couplings of the different aminoacid residues to be incorporated (from the N-terminal end to the C-terminal end in liquid phase, or from the C-terminal end to the N-terminal end in solid phase) wherein the N-terminal ends and the reactive side chains are previously blocked by conventional groups.

For solid phase synthesis the technique described by Merrifield (1965) may be used in particular.

J. ANTIBODIES

The TBC-1 polypeptide of SEQ ID No 5 can be used for the preparation of polyclonal or monoclonal antibodies.

The TBC-1 polypeptide expressed from a DNA sequence comprising at least one of the nucleic acid sequences of SEQ ID Nos 1, 3 and 4 may also be used to generate antibodies capable of specifically binding to the TBC-1 polypeptide of SEQ ID No 5.

Preferred antibodies according to the invention are prepared using TBC-1 peptide fragments that do not comprise the EVGYCQGL amino acid motif.

Other preferred antibodies of the invention are prepared using TBC-1 peptide fragments that do not comprise the TBC domain defined elsewhere in the specification.

The antibodies may be prepared from hybridomas according to the technique described by Kohler and Milstein in 1975. The polyclonal antibodies may be prepared by immunization of a mammal, especially a mouse or a rabbit, with a polypeptide according to the invention that is combined with

an adjuvant of immunity, and then by purifying of the specific antibodies contained in the serum of the immunized animal on a affinity chromatography column on which has previously been immobilized the polypeptide that has been used as the antigen.

The present invention also includes, chimeric single chain Fv antibody fragments (Martineau et al., 1998), antibody fragments obtained through phage display libraries (Ridder et al., 1995; Vaughan et al., 1995) and humanized antibodies (Reinmann et al., 1997; Leger et al., 1997).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

Consequently, the invention is also directed to a method for detecting specifically the presence of a TBC-1 polypeptide according to the invention in a biological sample, said method comprising the following steps :

a) bringing into contact the biological sample with an antibody according to the invention;

b) detecting the antigen-antibody complex formed.

The invention also concerns a diagnostic kit for detecting *in vitro* the presence of a TBC-1 polypeptide according to the present invention in a biological sample, wherein said kit comprises :

a) a polyclonal or monoclonal antibody as described above, optionally labeled;

b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent carrying optionally a label, or being able to be recognized itself by a labeled reagent, more particularly in the case when the above-mentioned monoclonal or polyclonal antibody is not labeled by itself.

K. SCREENING OF AGENTS ACTING AGAINST PROSTATE CANCER

In a further embodiment, the present invention also concerns a method for the screening of new agents, or candidate substances, acting against prostate cancer and which may be suitable for the treatment of a patient whose DNA comprises an allele of the *TBC-1* gene associated with prostate cancer, with an early onset of prostate cancer, or with the aggressiveness of prostate cancer tumors, or more generally with a modified or forthcoming expression of the *TBC-1* gene,

with a modified or forthcoming production of the *TBC-1* protein, or with the production of a modified *TBC-1* protein.

In a preferred embodiment, the invention relates to a method for the screening of candidate substances for prostate cancer treatment. The method comprises the following steps:

- providing a cell line, an organ, or a mammal expressing a *TBC-1* gene or a fragment thereof, preferably the regulatory region or the promoter region of the *TBC-1* gene.
- obtaining a candidate substance preferably a candidate substance capable of inhibiting the binding of a transcription factor to the *TBC-1* regulatory region,
- testing the ability of the candidate substance to decrease the symptoms of prostate cancer and/or to modulate the expression levels of *TBC-1*.

In some embodiments, the cell line, organ or mammal expresses a heterologous protein, the coding sequence of which is operably linked to the *TBC-1* regulatory or promoter sequence. In other embodiments, they express a *TBC-1* gene comprising alleles of one or more biallelic markers associated with prostate cancer an early onset of prostate cancer, or the aggressiveness of prostate cancer tumors, or a mutated *TBC-1* gene comprising a trait causing mutation determined using the above-noted method. In a further embodiment, a mouse expressing a *TBC-1* protein or a variant or a fragment thereof, more particularly by a nucleic acid comprising at least one of the biallelic markers according to the invention, can be used to screen agents acting against prostate cancer.

A candidate substance is a substance which can interact with or modulate, by binding or other intramolecular interactions, expression, stability, and function of *TBC-1*. Such substances may be potentially interesting for patients who are not responsive to existing drugs or develop side effects to them. Screening may be effected using either *in vitro* methods or *in vivo* methods.

Such methods can be carried out in numerous ways such as on transformed cells which express the considered alleles of the *TBC-1* gene, on tumors induced by said transformed cells, for example in mice, or on a TBC-1 protein encoded by the considered allelic variant of TBC-1.

Screening assays of the present invention generally involve determining the ability of a candidate substance to present a cytotoxic effect, to change the characteristics of transformed cells such as proliferative and invasive capacity, to affect the tumor growth, or to modify the expression level of TBC-1.

Typically, this method includes preparing transformed cells with different forms of *TBC-1* sequences containing particular alleles of one or more biallelic markers and/or trait causing mutations described above. This is followed by testing the cells expressing the TBC-1 with a candidate substance to determine the ability of the substance to present cytotoxic effect, to affect the characteristics of transformed cells, the tumor growth, or to modify the expression level of *TBC-1*.

Typical examples of such drug screening assays are provided below. It is to be understood that the parameters set forth in these examples can be modified by the skilled person without undue experimentation.

Methods for screening substances interacting with a TBC-1 polypeptide

For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to the TBC-1 protein or one of its fragments or variants or to modulate the expression of the polynucleotide coding for TBC-1 or a fragment or variant thereof.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of the TBC-1 protein is brought into contact with a purified TBC-1 protein, for example a purified recombinant TBC-1 protein produced by a recombinant cell host as described hereinbefore, in order to form a complex between the TBC-1 protein and the putative ligand molecule to be tested.

A. Candidate ligands obtained form random peptide libraries

In a particular embodiment of the screening method, the putative ligand is the expression product of a DNA insert contained in a phage vector (Parmley and Smith, 1988). Specifically, random peptide phages libraries are used. The random DNA inserts encode peptides of 8 to 20 aminoacids in length (Oldenburg K.R. et al., 1992,.; Valadon P., et al., 1996; Lucas A.H., 1994; Westerink M.A.J., 1995; Castagnoli L. et al., 1991). According to this particular embodiment, the recombinant phages expressing a protein that binds to the immobilized TBC-1 protein are retained and the complex formed between the TBC-1 protein and the recombinant phage may be subsequently immunoprecipitated by a polyclonal or a monoclonal antibody directed against the TBC-1 protein.

Once the ligand library in recombinant phages has been constructed, the phage population is brought into contact with the immobilized TBC-1 protein. Then the preparation of complexes is washed in order to remove the non-specifically bound recombinant phages. The phages that bind specifically to the TBC-1 protein are then eluted by a buffer (acid pH) or immunoprecipitated by the anti-TBC-1monoclonal antibody produced by a hybridoma, and this phage population is subsequently amplified by an over-infection of bacteria (for example E. coli). The selection step may be repeated several times, preferably 2-4 times, in order to select the more specific recombinant phage clones. The last step consists in characterizing the peptide produced by the selected recombinant phage clones either by expression in infected bacteria and isolation, expressing the phage insert in another host-vector system, or sequencing the insert contained in the selected recombinant phages.

B. Candidate ligands obtained through a two-hybrid screening assay.

The yeast two-hybrid system is designed to study protein-protein interactions *in vivo* (Fields and Song, 1989), and relies upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein. This technique is also described in US Patent N° US 5,667,973 and US Patent N° 5,283,173 (Fields et al.) the technical teachings of both patents being herein incorporated by reference.

The general procedure of library screening by the two-hybrid assay may be performed as described by Harper et al. (Harper JW et al., 1993) or as described by Cho et al. (1998) or also Fromont-Racine et al. (1997).

The bait protein or polypeptide consists of a TBC-1 polypeptide or a fragment or variant thereof.

More precisely, the nucleotide sequence encoding the TBC-1 polypeptide or a fragment or variant thereof is fused to a polynucleotide encoding the DNA binding domain of the GAL4 protein, the fused nucleotide sequence being inserted in a suitable expression vector, for example pAS2 or pM3.

Then, a human cDNA library is constructed in a specially designed vector, such that the human cDNA insert is fused to a nucleotide sequence in the vector that encodes the transcriptional domain of the GAL4 protein. Preferably, the vector used is the pACT vector. The polypeptides encoded by the nucleotide inserts of the human cDNA library are termed "pray" polypeptides.

A third vector contains a detectable marker gene, such as beta galactosidase gene or CAT gene that is placed under the control of a regulation sequence that is responsive to the binding of a complete Gal4 protein containing both the transcriptional activation domain and the DNA binding domain. For example, the vector pG5EC may be used.

Two different yeast strains are also used. As an illustrative but non limiting example the two different yeast strains may be the following :

- Y190, the phenotype of which is (*MATa, Leu2-3, 112 ura3-12, trp1-901, his3-D200, ade2-101, gal4Dgal180D URA3 GAL-LacZ, LYS GAL-HIS3, cyh*^{r});
- Y187, the phenotype of which is (*MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3, -112 URA3 GAL-lacZmet*$^-$), which is the opposite mating type of Y190.

Briefly, 20 µg of pAS2/TBC-1 and 20 µg of pACT-cDNA library are co-transformed into yeast strain Y190. The transformants are selected for growth on minimal media lacking histidine, leucine and tryptophan, but containing the histidine synthesis inhibitor 3-AT (50 mM). Positive colonies are screened for beta galactosidase by filter lift assay. The double positive colonies (*His*$^+$, *beta-gal*$^+$) are then grown on plates lacking histidine, leucine, but containing tryptophan and cycloheximide (10 mg/ml) to select for loss of pAS2/TBC-1 plasmids but retention of pACT-cDNA library plasmids. The resulting Y190 strains are mated with Y187 strains expressing TBC-1 or non-related control proteins; such as cyclophilin B, lamin, or SNF1, as *Gal4* fusions as described by Harper et al. (1993) and by Bram et al. (1993), and screened for beta galactosidase by filter lift assay. Yeast clones that are *beta gal-* after mating with the control *Gal4* fusions are considered false positives.

In another embodiment of the two-hybrid method according to the invention, the interaction between TBC-1 or a fragment or variant thereof with cellular proteins may be assessed using the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech).). As described in the manual accompanying the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech), the disclosure of which is incorporated herein by reference, nucleic acids encoding the TBC-1 protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. A desired cDNA, preferably human cDNA, is inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into the yeast cells and the yeast cells are plated on selection medium which selects

for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay are those in which an interaction between TBC-1 and the protein or peptide encoded by the initially selected cDNA insert has taken place.

Method for screening ligands that modulate the expression of the *TBC-1* gene.

Another subject of the present invention is a method for screening molecules that modulate the expression of the TBC-1 protein. Such a screening method comprises the steps of :

a) cultivating a prokaryotic or an eukaryotic cell that has been transfected with a nucleotide sequence encoding the TBC-1 protein, operably linked to a *TBC-1* 5'-regulatory sequence;

b) bringing into contact the cultivated cell with a molecule to be tested;

c) quantifying the expression of the TBC-1 protein.

Using DNA recombination techniques well known by the one skill in the art, the TBC-1 protein encoding DNA sequence is inserted into an expression vector, downstream from a *TBC-1* 5'-regulatory sequence that contains a *TBC-1* promoter sequence. As an illustrative example, the promoter sequence of the *TBC-1* gene is contained in the nucleic acid of SEQ ID No 2.

The quantification of the expression of the TBC-1 protein may be realized either at the mRNA level or at the protein level. In the latter case, polyclonal or monoclonal antibodies may be used to quantify the amounts of the TBC-1 protein that have been produced, for example in an ELISA or a RIA assay.

In a preferred embodiment, the quantification of the *TBC-1* mRNAs is realized by a quantitative PCR amplification of the cDNAs obtained by a reverse transcription of the total mRNA of the cultivated *TBC-1*-transfected host cell, using a pair of primers specific for *TBC-1*.

Expression levels and patterns of *TBC-1* may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277, the entire contents of which are incorporated herein by reference. Briefly, the *TBC-1* cDNA or the *TBC-1* genomic DNA described above, or fragments thereof, is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the *TBC-1* insert comprises at least 100 or more consecutive nucleotides of the genomic DNA sequence or the cDNA sequences, particularly those comprising one of the nuceotide sequences

of SEQ ID Nos 3, 4 and 6-8 or those encoding a mutated TBC-1. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

Quantitative analysis of *TBC-1* gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of a plurality of nucleic acids of sufficient length to permit specific detection of expression of mRNAs capable of hybridizing thereto. For example, the arrays may contain a plurality of nucleic acids derived from genes whose expression levels are to be assessed. The arrays may include the *TBC-1* genomic DNA, the *TBC-1* cDNA sequences or the sequences complementary thereto or fragments thereof, particularly those comprising at least one of the biallelic markers according the present invention, preferably at least one of SEQ ID 7-8 or those comprising a trait causing mutation. Preferably, the fragments are at least 15 nucleotides in length. In other embodiments, the fragments are at least 25 nucleotides in length. In some embodiments, the fragments are at least 50 nucleotides in length. More preferably, the fragments are at least 100 nucleotides in length. In another preferred embodiment, the fragments are more than 100 nucleotides in length. In some embodiments the fragments may be more than 500 nucleotides in length.

For example, quantitative analysis of *TBC-1* gene expression may be performed with a complementary DNA microarray as described by Schena et al. (1995). Full length *TBC-1* cDNAs or fragments thereof are amplified by PCR and arrayed from a 96-well microtiter plate onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm^2 microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of *TBC-1* gene expression may also be performed with full length *TBC-1* cDNAs or fragments thereof in complementary DNA arrays as described by Pietu et al. (1996). The full length *TBC-1* cDNA or fragments thereof is PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis using the *TBC-1* genomic DNA, the *TBC-1* cDNAs, or fragments thereof can be done through high density nucleotide arrays or chips as described by Lockhart et al. (1996) and Sosnowsky et al. (1997). Oligonucleotides of 15-50 nucleotides from the sequences of the *TBC-1* genomic DNA, the *TBC-1* cDNA sequences particularly those comprising at least one of biallelic markers according the present invention, preferably at least one of SEQ ID No 7-8 or those comprising the trait causing mutation, or the sequences complementary thereto, are synthesized directly on the chip (Lockhart et al., supra) or synthesized and then addressed to the chip (Sosnowski et al., supra). Preferably, the oligonucleotides are about 20 nucleotides in length.

TBC-1 cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart et al., supra and application of different electric fields (Sosnowsky et al., 1997)., the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of *TBC-1* mRNAs.

Screening using transgenic animals

In vivo methods can utilize transgenic animals for drug screening. Nucleic acids including at least one of the biallelic polymorphisms of interest can be used to generate genetically modified non-human animals or to generate site specific gene modifications in cell lines. The term "transgenic" is intended to encompass genetically modified animals having a deletion or other knock-out of *TBC-1* gene activity, having an exogenous *TBC-1* gene that is stably transmitted in the host cells, or having an exogenous *TBC-1* promoter operably linked to a reporter gene. Transgenic animals may be made through homologous recombination, where the *TBC-1* locus is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include for example plasmids, retroviruses and other animal viruses, and YACs. Of interest are transgenic mammals e.g. cows, pigs, goats, horses, and particularly rodents such as rats and mice. Transgenic animals allow to study both efficacy and toxicity of the candidate drug.

Methods for inhibiting the expression of a *TBC-1* gene

Other therapeutic compositions according to the present invention comprise advantageously an oligonucleotide fragment of the nucleic sequence of *TBC-1* as an antisense tool that inhibits the expression of the corresponding *TBC-1* gene. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al. (1995).

Preferably, the antisense tools are choosen among the polynucleotides (15-200 bp long) that are complementary to the 5'end of the *TBC-1* mRNA. In another embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targetted gene are used.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of *TBC-1* that contains the translation initiation codon ATG

The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. They comprise a nucleotide sequence complementary to the targeted sequence of the PTCA-1 genomic DNA, the sequence of which can be determined using one of the detection methods of the present invention. The targeted DNA or RNA sequence preferably comprises at least one of the biallelic markers according to the present invention, particularly a nucleotide

sequence selected from the group consisting of SEQ ID 7-14, or comprises a trait causing mutation. In a preferred embodiment, the antisense oligonucleotide are able to hybridize with at least one of the splicing sites of the targeted *TBC-1* gene, with the 3'UTR of the 5'UTR, with exon 6bis, or with an exonic region comprising at least one of the biallelic markers of the present invention or comprising a trait causing mutation. The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the *TBC-1* mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green et al., (1986) and Izant and Weintraub, (1984), the disclosures of which are incorporated herein by reference.

In some strategies, antisense molecules are obtained by reversing the orientation of the *TBC-1* coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using in vitro transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of *TBC-1* antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

Alternatively, suitable antisense strategies are those described by Rossi et al. (1991), in the International Applications Nos. WO 94/23026, WO 95/04141, WO 92/18522 and in the European Patent Application No. EP 0 572 287 A2

An alternative to the antisense technology that is used according to the present invention consists in using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely « hammerhead ribozymes »). Briefly, the simplified cycle of a hammerhead ribozyme consists of (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage products, which gives rise to another catalytic cycle. Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense ribozymes according to the present invention are prepared as

described by Sczakiel et al. (1995), the specific preparation procedures being referred to in said article being herein incorporated by reference.

EXAMPLES

EXAMPLE 1 : Analysis of the first mRNA encoding a TBC-1 polypeptide synthesized by the cells.

TBC-1 cDNA was obtained as follows : 4µl of ethanol suspension containing 1 mg of human prostate total RNA (Clontech laboratories, Inc., Palo Alto, USA; Catalogue N. 64038-1) was centrifuged, and the resulting pellet was air dried for 30 minutes at room temperature.

First strand cDNA synthesis was performed using the AdvantageTM RT-for- PCR kit (Clontech laboratories Inc., catalogue N. K1402-1). 1 µl of 20 mM solution of a specific oligo dT primer was added to 12.5 µl of RNA solution in water, heated at 74°C for 2.5 min and rapidly quenched in an ice bath. 10 µl of 5 x RT buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 2.5 µl of dNTP mix (10 mM each), 1.25 µl of human recombinant placental RNA inhibitor were mixed with 1 ml of MMLV reverse transcriptase (200 units). 6.5 µl of this solution were added to RNA-primer mix and incubated at 42°C for one hour. 80 µl of water were added and the solution was incubated at 94°C for 5 minutes.

5µl of the resulting solution were used in a Long Range PCR reaction with hot start, in 50 µl final volume, using 2 units of rtTHXL, 20 pmol/µl of each of 5'-TGACCACCATGCCCATGCT-3' (SEQ ID No 13) and 5'-GCATTTATTCACGTCCACGCC-3' (SEQ ID No 14) primers with 35 cycles of elongation for 6 minutes at 67°C in thermocycler.

The amplification products corresponding to both cDNA strands were partially sequenced in order to ensure the specificity of the amplification reaction.

Results of Nothern blot analysis of prostate mRNAs supported the existence of the first *TBC-1* cDNA having about 4 kb in length, which is the nucleotide sequence of SEQ ID No 3.

Example 2 : Detection of *TBC-1* biallelic markers: DNA extraction

Donors were unrelated and healthy. They presented a sufficient diversity for being representative of a French heterogeneous population. The DNA from 100 individuals was extracted and tested for the detection of the biallelic markers.

30 ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume : 10 mM Tris pH7.6; 5 mM $MgCl_2$; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution.

The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M
- 200 µl SDS 10%
- 500 µl K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50 µg/ml DNA). To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below.

The pool was constituted by mixing equivalent quantities of DNA from each individual.

Example 3 : Detection of the biallelic markers: amplification of genomic DNA by PCR

The amplification of specific genomic sequences of the DNA samples of example 2 was carried out on the pool of DNA obtained previously. In addition, 50 individual samples were similarly amplified.

PCR assays were performed using the following protocol:

Final volume	25 µl

DNA	2 ng/µl
$MgCl_2$	2 mM
dNTP (each)	200 µM
primer (each)	2.9 ng/µl
Ampli Taq Gold DNA polymerase	0.05 unit/µl
PCR buffer (10x = 0.1 M TrisHCl pH8.3 0.5M KCl	1x

Each pair of first primers was designed using the sequence information of the *TBC-1* gene disclosed herein and the OSP software (Hillier & Green, 1991). This first pair of primers was about 20 nucleotides in length and had the sequences disclosed in Table 1 in the columns labeled PU and RP.

Table 1

	Amplified region of *TBC-1* gene	PU	RP
1	Intron 1	SEQ ID No 9	SEQ ID No 10

Preferably, the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification which was useful for sequencing.

Primers PU contain the following additional PU 5' sequence : TGTAAAACGACGGCCAGT; primers RP contain the following RP 5' sequence : CAGGAAACAGCTATGACC. The primer containing the additional PU 5' sequence is listed in SEQ ID No 9. The primer containing the additional RP 5' sequence is listed in SEQ ID No 10.

The synthesis of these primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

DNA amplification was performed on a Genius II thermocycler. After heating at 95°C for 10 min, 40 cycles were performed. Each cycle comprised: 30 sec at 95°C, 54°C for 1 min, and 30 sec at 72°C. For final elongation, 10 min at 72°C ended the amplification. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

Example 4 : Detection of the biallelic markers: sequencing of amplified genomic DNA and identification of polymorphisms.

The sequencing of the amplified DNA obtained in example 3 was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis [ABI Prism DNA Sequencing Analysis software (2.1.2 version) and the above mentioned proprietary "Trace" basecaller].

The sequence data were further evaluated using the above mentioned polymorphism analysis software designed to detect the presence of biallelic markers among the pooled amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position as described previously.

One fragment of amplification was analyzed. In this segment, 1 biallelic marker was detected. The localization of the biallelic marker is as shown in Table 2.

Table 2

Amplicon	Marker Name	Localization in *TBC-1* gene	Polymorphism	Major allele	Minor allele
1	99-430-352	Intron 1	C/T	SEQ ID No 7	SEQ ID No 8

Example 5 : Validation of the polymorphisms through microsequencing

The biallelic markers identified in example 4 were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was carried out for each individual DNA sample described in Example 2.

Amplification from genomic DNA of individuals was performed by PCR as described above for the detection of the biallelic markers with the same set of PCR primers (Table 1).

The preferred primers used in microsequencing were about 23 nucleotides in length and hybridized just upstream of the considered polymorphic base. According to the invention, the primers used in microsequencing are detailed in Table 3.

Table 3

Marker Name	PU Microsequencing primer	RP microsequencing primer
99-430-352	SEQ ID No 11	SEQ ID No 12

The microsequencing reaction was performed as follows :

After purification of the amplification products, the microsequencing reaction mixture was prepared by adding, in a 20µl final volume: 10 pmol microsequencing oligonucleotide, 1 U Thermosequenase (Amersham E79000G), 1.25 µl Thermosequenase buffer (260 mM Tris HCl pH 9.5, 65 mM $MgCl_2$), and the two appropriate fluorescent ddNTPs (Perkin Elmer, Dye Terminator Set 401095) complementary to the nucleotides at the polymorphic site of each biallelic marker tested, following the manufacturer's recommendations. After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The unincorporated dye terminators were then

removed by ethanol precipitation. Samples were finally resuspended in formamide-EDTA loading buffer and heated for 2 min at 95°C before being loaded on a polyacrylamide sequencing gel. The data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Following gel analysis, data were automatically processed with software that allows the determination of the alleles of biallelic markers present in each amplified fragment.

The software evaluates such factors as whether the intensities of the signals resulting from the above microsequencing procedures are weak, normal, or saturated, or whether the signals are ambiguous. In addition, the software identifies significant peaks (according to shape and height criteria). Among the significant peaks, peaks corresponding to the targeted site are identified based on their position. When two significant peaks are detected for the same position, each sample is categorized classification as homozygous or heterozygous type based on the height ratio.

Example 6 : Association study between prostate cancer and the biallelic marker of the *TBC-1* gene: collection of DNA samples from affected and non-affected individuals

Affected population :

The positive trait in this association study was prostate cancer. Prostate cancer patients were recruited according to a combination of clinical, histological and biological inclusion criteria. Clinical criteria can include rectal examination and prostate biopsies. Biological criteria can include PSA assays. The affected individuals were recorded as familial forms when at least two persons affected by prostate cancer have been diagnosed in the family. Remaining cases were classified as non-familial informative cases (at least two sibs of the case both aged over 50 years old are unaffected), or non-familial uniformative cases (no information about sibs over 50 years old is available). Cases were also separated following the criteria of diagnosis age : early onset prostate cancer (under 65 years old) and late onset prostate cancer (65 years old or more).

The different populations included in the association study of this patent are characterized in more detail in Table 4.

Table 4

POP	AFFECTED POPULATIONS		UNAFFECTED POPULATIONS
M2S2	- affected : unrelated individuals		- unaffected : healthy ind. of + 65 years old

M3S2	- affected : individuals diagnosed before 65 years old			- unaffected : healthy ind. of + 65 years old
M4S2	- affected : individuals diagnosed after 65 years old			- unaffected : healthy ind. of + 65 years old
M5S2	- affected : sporadic cases			- unaffected : healthy ind. of + 65 years old
M6S2	- affected : familial cases			- unaffected : healthy ind. of + 65 years old

Several populations were studied among the affected group, which were the followings :

1) Unaffected population :

Control individuals included in this study were checked for both the absence of all clinical and biological criteria defining the presence or the risk of prostate cancer (PSA < 4 ng/ml) (WO 96/21042), and for their age (aged 65 years old or more). All unaffected individuals included in the statistical analysis of this sample were unrelated. The unaffected population has a size of 127 individuals.

2) Affected populations :

a) The class M2S2 consisted of 342 unrelated individuals affected by prostate cancer;

b) The class M3S2 consisted of 136 unrelated individuals affected by prostate cancer and which have been diagnosed after the age of 65 years old. The patients of this class were included in the class M2S2;

c) The class M4S2 consisted in 206 unrelated individuals affected by prostate cancer and which had been diagnosed before the age of 65 years old. The patients of this class were included in the class M2S2.

d) The class M5S2 consisted in 176 prostate cancer sporadic cases.

e) The class M6S2 consisted in 166 unrelated individuals belonging to families wherein at least two members had been affected by prostate cancer.

EXAMPLE 7 : Association study between prostate cancer and the biallelic marker of the *TBC-1* gene: Genotyping of affected and control individuals.

The general strategy to perform the association studies was to individually scan the DNA samples from all individuals in each of the populations described above in order to establish the allele frequencies of biallelic marker 99-430-352 in each of these populations.

Allelic frequencies of the above-described biallelic marker in each population were determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual. Genomic PCR and microsequencing were

performed as detailed above in examples 3 and 5 using the described PCR and microsequencing primers. The results are depicted in Table 5.

Table 5

MARKER	AFFECTED POPULATION				UNAFFECTED POP.				POP
	Pop. Size	C/C	T/T	C/T	Pop. Size	C/C	T/T	C/T	
99-430-352	342	59	7	33	127	48	9	41	M2S2
99-430-352	136	61	3	34	127	48	9	41	M3S2
99-430-352	206	57	9	33	127	48	9	41	M4S2
99-430-352	176	66	3	29	127	48	9	41	M5S2
99-430-352	166	51	10	37	127	48	9	41	M6S2

The column labeled "Pop. Size" indicates the number of individuals for which the DNA was genotyped using the microsequencing primer having the nucleotide sequence of SEQ ID No 11.

The columns labeled C/C, T/T and C/T correspond to the respective frequencies of the different allele polymorphisms for biallelic marker 99-430-352 in the diploid genome of the tested individuals.

The column POP depicts the different populations under consideration that have been previously described in Example 5.

EXAMPLE 8 : Association study between prostate cancer and the biallelic marker of the *TBC-1* gene.

Table 6 shows, for each of the populations studied, the frequency of each allele and the statistical significance of an imbalance in the frequency of the major allele of the biallelic marker 99-430-352 between the unaffected population and each of the affected populations described in Example 5.

Table 6

MARKERS	AFFECTED POPULATIONS			UNAFFECTED POPULATIONS			diff (chi2 - pvalue)	POP
	Pop. Size	C	T	Pop. Size	C	T		
[illegible]	342	75.88	24.12	127	69.69	30.31	6.2 (3.71 - 5.13e-02)	**M2S2**
[illegible]	136	79.04	20.96	127	69.69	30.31	9.4 (6.06 - 1.35e-02)	**M3S2**
[illegible]	206	73.79	26.21	127	69.69	30.31	4.1 (1.32 - 2.37e-01)	**M4S2**
[illegible]	176	81.25	18.75	127	69.69	30.31	**11.6(10.95 - 9.11e-04)**	**M5S2**
[illegible]	166	70.18	29.82	127	69.69	30.31	0.5 (0.02 - 7.52e-01)	**M6S2**

The column labeled "Pop. Size" indicates the number of individuals for which the DNA was genotyped using the microsequencing primer having the nucleotide sequence of SEQ ID No 11.

The column labeled "C" or "T" show the relative allele frequency of biallelic marker 99-430-352, the percentage value being expressed for a haploid genome and calculated as follows :

a) Frequency (F) of the 99-430-352 biallelic marker allele bearing C as the polymorphic base for the M2S2 population :

$F_{C\text{-}M2S2} = [(C/C_{M2S2} \times \text{Pop. Size}_{M2S2} \times 2) + (C/T_{M2S2} \times \text{Pop. Size}_{M2S2})] / [\text{Pop. Size}_{M2S2} \times 2]$.

b) Frequency (F) of the 99-430-352 biallelic marker allele bearing T as the polymorphic base for the M2S2 population :

$F_{T\text{-}M2S2} = [(T/T_{M2S2} \times \text{Pop. Size}_{M2S2} \times 2) + (C/T_{M2S2} \times \text{Pop. Size}_{M2S2})] / [\text{Pop. Size}_{M2S2} \times 2]$.

The column "diff (chi2 – pvalue)" show the following values :

a) "diff" is the result of the difference in the allele frequency of the more frequent allele, which is allele "C", between the affected and unaffected individuals, as detailed in Table 6.

b) "chi2" is the result of the statistical Chi2 test when taking into account the "diff" value and the size of both the affected and unaffected populations under consideration.

c) "p value" is calculated from the "chi2" value, with a degree of freedom value of 1.

The higher the difference of frequency is, the higher the chi2 value is and the lower the p value is, the stronger the association between the marker and the trait is.

It has been considered by the inventors that a pvalue of 5% (5×10^{-2}) is statistically significant.

From the data shown in Table 6, it can be observed that there is a high chi2 value (10.95) and a very low p value (9.11×10^{-4}) when testing the M5S2 population, which corresponds to the sporadic cases. It can be observed that the pvalue is far lower than 5×10^{-2} and thus that the difference in allele frequency between the healthy population and the prostate cancer sporadic cases is highly significant.

It can then be stated that a very significant association exists between the 99-430-352 biallelic marker and prostate cancer, and more particularly with sporadic prostate cancer cases.

While the preferred embodiments of the invention have been illustrated end described, it will be appreciated that various changes can be made therein by the one skilled in the art without departing from the spirit and scope of the invention.

References

Bram RJ et al., 1993, Mol. Cell Biol., **13** : 4760-4769.

Castagnoli L. et al. (Felici F.), 1991, J. Mol. Biol., **222**:301-310.

Chai H. et al., 1993, Biotechnol. Appl. Biochem., **18**:259-273

Cherif et al., 1990, Proc. Natl. Acad. Sci. USA, **87** : 6639-6643.

Cho RJ et al., 1998, Proc. Natl. Acad. Sci. USA, **95**(7) : 3752-3757.

Chumakov I. et al., 1995, Nature, **377**(6547 Suppl): 175-297.

Dib et al., 1996, Nature, **380**: III-V.

Ellis NA,1997 Curr.Op.Genet.Dev., **7** : 354-363

Feldman and Steg, 1996, Medecine/Sciences, synthese, **12**:47-55

Fields and Song, 1989, Nature, Vol. 340 : 245-246.

Fishel R & Wilson T. 1997, Curr.Op.Genet.Dev.7: 105-113

Flotte et al., 1992, Am. J. Respir. Cell Mol. Biol., 7 : 349-356.

Fromont-Racine M. et al., 1997, Nature Genetics, **16**(3) : 277-282.

Fuller S.A. et al., 1996, Immunology in Current Protocols in Molecular Biology, Ausubel et al. Eds, John Wiley & Sons, Inc., USA

Green et al., *Ann. Rev. Biochem.* **55**:569-597 (1986)

Grompe, M. et al., *Proc. Natl. Acad. Sci. U.S.A* 1989; **86**:5855-5892

Grompe, M. *Nature Genetics* 1993; **5**:111-117

Haber D & Harlow E, 1997, Nature Genet. **16**:320-322.

Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS, *Nat Genet* 1996;**14**(4):441-447

Harju L, et al., *Clin Chem* 1993;**39**(11Pt 1):2282-2287

Harper JW et al., 1993, Cell, **75** : 805-816.

Harper JW et al., 1993, Cell, Vol. 75 : 805-816.

Harris H et al.,1969,Nature 223:363-368.

Hillier L. and Green P. *Methods Appl.*, 1991, **1**: 124-8.

Houbenweyl, 1974, in Meuthode der Organischen Chemie, E. Wunsch Ed., Volume 15-I et 15-II, Thieme, Stuttgart.

Huygen et al., 1996, Nature Medicine, **2**(8):893-898

Izant and Weintraub, *Cell* **36**:1007-1015 (1984)

Julan et al., 1992, J. Gen. Virol., **73** : 3251 – 3255.

Koch Y., 1977, Biochem. Biophys. Res. Commun., **74**:488-491

Kohler G. and Milstein C., 1975, Nature, **256** : 495.

Kozal MJ, Shah N, Shen N, Yang R, Fucini R, Merigan TC, Richman DD, Morris D, Hubbell E, Chee M, Gingeras TR, *Nat Med* 1996;**2**(7):753-759

Leger OJ, et al., 1997, Hum Antibodies, **8**(1): 3-16

Lenhard T. et al., 1996, Gene, **169**:187-190

Livak KJ, and Hainer JW., 1994, Hum Mutat., **3**(4): 379-385.

Lockhart et al. *Nature Biotechnology* **14**: 1675-1680, 1996

Lucas A.H., 1994, In : Development and Clinical Uses of Haempophilus b Conjugate.

Mansour SL et al., 1988, Nature, **336** : 348-352.

Manz et al., *Adv in Chromatogr* 1993; **33**:1-66

Mc Whorter W.P., Hernandez A.D., Meikle A.W. et al. A screening study of prostate cancer in high risk families. J Urol 1992;148:826-828.

McLaughlin et al., 1989, J. Virol., **62** : 1963 – 1973.

Martineau P, Jones P, Winter G, 1998, J Mol Biol, **280**(1):117-127

Merrifield RB, 1965, Nature, **207**(996): 522-523.

Merrifield RB., 1965, Science, **150**(693): 178-185.

Muzyczka et al., 1992, Cuur. Topics in Micro. and Immunol., **158** : 97-129.

Neda et al., 1991, J. Biol. Chem., **266** : 14143 – 14146.

O'Reilly et al., 1992, Baculovirus expression vectors : a Laboratory Manual. W.H. Freeman and Co., New York

Ohno et al., 1994, Sciences, **265**:781-784

Oldenburg K.R. et al., 1992, Proc. Natl. Acad. Sci., **89**:5393-5397.

Orita et al., *Proc. Natl. Acad. Sci. U.S.A.* 1989;**86**: 2776-2770

Parmley and Smith, Gene, 1988, **73**:305-318.

Pastinen et al., *Genome Research* 1997; **7**:606-614

PCR Methods and Applications", 1991, Cold Spring Harbor Laboratory Press.

Pietu et al. *Genome Research* **6**:492-503, 1996

Porath J et al., 1975, Nature, **258**(5536) : 598-599.

Reimann KA, et al., 1997, AIDS Res Hum Retroviruses. **13**(11): 933-943

Ridder R, Schmitz R, Legay F, Gram H, 1995, Biotechnology (N Y), **13**(3):255-260

Rossi et al., *Pharmacol. Ther.* **50**:245-254, (1991)

Roth J.A. et al., 1996, Nature Medicine, **2**(9):985-991

Rougeot, C. et al.,. *Eur. J. Biochem.* **219** (3): 765-773, 1994

Roux et al., 1989, Proc. Natl Acad. Sci. USA, **86** : 9079 – 9083.

Sambrook, J. Fritsch, E. F., and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York.

Samulski et al., 1989, J. Virol., **63** : 3822-3828.

Sanchez-Pescador R., 1988, J. Clin. Microbiol., **26**(10):1934-1938

Sandou et al., 1994, Science, **265** : 1875-1878.

Schena et al. *Science* **270**:467-470, 1995

Sczakiel G. et al., 1995, Trends Microbiol., 1995, 3(6):213-217

Sheffield, V.C. et al, *Proc. Natl. Acad. Sci. U.S.A* 1991; **49**:699-706

Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW, *Nat Genet* 1996;**14**(4):450-456

Smith et al., 1983, Mol. Cell. Biol., 3:2156-2165.

Sosnowski RG, et al., *Proc Natl Acad Sci U S A* 1997;**94**:1119-1123

Sosnowski RG, et al., *Proc Natl Acad Sci U S A* 1997;**94**:1119-1123

Steinberg G.D., Carter B.S., Beaty T.H. et al. Family history and the risk of prostate cancer, The prostate 1990;17,337-347.

Strittmatter WJ, et al., *Proc Natl Acad Sci U S A* 1993 ;**90**:1977-1981

Syvanen AC, et al., 1994, Hum Mutat., 3(3): 172-179.

Tacson et al., 1996, Nature Medicine, 2(8):888-892.

Urdea M.S., 1988, Nucleic Acids Research, **11**: 4937-4957

Urdea MS et al., 1991, Nucleic Acids Symp Ser., **24**: 197-200.

Valadon P., et al., 1996, J. Mol. Biol., Vol. 261:11-22.

Vaughan TJ, et al., 1996, Nat Biotechnol. **14**(3): 309-314

Vlasak R. et al., 1983, Eur. J. Biochem., **135**:123-126

Wabiko et al., 1986, DNA, 5(4):305-314.

Westerink M.A.J., 1995, Proc. Natl. Acad. Sci., **92**:4021-4025.

Wilson R. et al., 1994, Nature, **368**(6466) : 32-38.

Zhang SD et al., 1996, Genes and development, **10** : 1108-1119.

What is claimed is :

1. A purified or isolated nucleic acid encoding a TBC-1 protein having the amino acid sequence of SEQ ID No 5, or a peptide fragment or variant thereof.

2. A purified or isolated nucleic acid comprising a nucleotide sequence of SEQ ID No 1, or a sequence complementary thereto.

3. A purified or isolated nucleic acid comprising a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4, or fragments or variants thereof.

4. A purified or isolated nucleic acid having at least 95% of nucleotide identity with any one of the nucleotide sequences of SEQ ID Nos 3 and 4, or fragments thereof.

5. A purified or isolated nucleic acid according to claim 4, wherein this nucleic acid comprises the nucleotide sequence of SEQ ID No 6

6. A purified or isolated nucleic acid comprising the nucleotide sequence beginning at the nucleotide in position 171 and ending at the nucleotide in position 3728 of the polynucleotide of SEQ ID No 3.

7. A purified or isolated nucleic acid comprising the nucleotide sequence of SEQ ID No 2, or a biologically active fragment or variant thereof.

8. The nucleic acid of claim 7, wherein said nucleic acid further comprises a purified or isolated nucleic acid comprising a nucleotide sequence located between the nucleotide at position 1 and the nucleotide at position 170 of SEQ ID No 3, or a biologically active fragment or variant thereof.

9. The nucleic acid of claim 7, wherein said nucleic acid further comprises a purified or isolated nucleic comprising a nucleotide sequence located between the nucleotide at position 1 and the nucleotide at position 175 of SEQ ID No 4, or a biologically active fragment or variant thereof.

10. A purified or isolated nucleic acid comprising :

a) a nucleic acid comprising a regulatory nucleotide sequence of SEQ ID No 2, or a biologically active fragment or variant thereof;

b) a polynucleotide encoding a desired polypeptide or nucleic acid, operably linked to the nucleic acid comprising a regulatory nucleotide sequence of SEQ ID No 2, or its biologically active fragment or variant;

c) Optionally, a nucleic acid comprising a 3'-UTR regulatory polynucleotide, preferably a 3'-UTR polynucleotide of the *TBC-1* gene.

11. The nucleic acid of claim 10, wherein the polynucleotide of step b) encodes a sense or an antisense RNA molecule.

12. A purified or isolated nucleic acid useful as an amplification primer or as a probe, wherein this nucleic acid comprises a polynucleotide selected from the group consisting of the nucleotide sequences complementary to any sequence of a strand of the *TBC-1* gene and the nucleotide sequences complementary to any one of the SEQ ID Nos 2-4, 7-8 and 13-14.

13. A nucleic acid probe or primer comprising at least 20 consecutive nucleotides of the nucleotide sequence of SEQ ID No 1.

14. A nucleic acid probe or primer comprising at least 8 consecutive nucleotides of a polynucleotide selected from the group consisting of :

a) the nucleotide sequence beginning at the nucleotide in position 2077 and ending at the nucleotide in position 12290 of the nucleotide sequence of SEQ ID No 1;

b) the nucleotide sequence beginning at the nucleotide in position 12373 and ending at the nucleotide in position 12738 of the nucleotide sequence of SEQ ID No 1.

15. A nucleic acid probe or primer comprising at least 8 consecutive nucleotides of a polynucleotide of SEQ ID No 2.

16. A nucleic acid probe or primer comprising at least 20 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4.

17. A nucleic acid probe or primer selected from the group consisting of the nucleotide sequences of SEQ ID Nos 9-14.

18. A purified or isolated nucleic acid comprising a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 7-8.

19. A purified or isolated biallelic marker , wherein said biallelic marker is from the sequence of the *TBC-1* gene.

20. A purified or isolated biallelic marker according to claim 17, wherein said biallelic marker is associated with prostate cancer.

21. A purified or isolated nucleic acid selected from the group consisting of SEQ ID No 7-8 or a variant or fragment thereof, said fragment comprising at least 8 consecutive nucleotides of said nucleic acid and including the polymorphic base thereof.

22. A method for the identification and characterization of a biallelic marker in the genomic region harboring the *TBC-1* gene, said method comprising :

- providing a plurality of primer sequences capable of amplifying portions of the genomic region containing the *TBC-1* gene, and in particular portions of the polynucleotide of SEQ ID No 1;
- amplifying portions of the genomic region containing the *TBC-1* gene from a plurality of individuals using said primers to obtain a plurality of amplicons; and
- sequencing said plurality of amplicons to identify biallelic markers in the genomic region harboring the *TBC-1* gene.

23. A method for the amplification of the *TBC-1* gene or a fragment or a variant thereof in a test sample, said method comprising the steps of :

c) contacting a test sample suspected of containing the targeted *TBC-1* gene sequence or portion thereof with amplification reaction reagents comprising a pair of amplification primers located on either side of the *TBC-1* region to be amplified, and

d) detecting the amplification products.

24. The method according to claim 23, wherein the amplification primers are selected from the group consisting of SEQ ID Nos 9-10.

25. The method according to claim 23, wherein the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to a region of the *TBC-1* gene.

26. A kit for the amplification of a nucleotide sequence contained in the *TBC-1* gene, wherein said kit comprises :

a) A pair of oligonucleotide primers located on either side of the *TBC-1* region to be amplified;

b) Optionally, the reagents necessary for performing the amplification reaction.

27. A method for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 1-4 and 7-8 in a sample, said method comprising the following steps of :

a) bringing into contact a nucleic acid probe or a plurality of nucleic acid probes which can hybridize to a nucleotide sequence included in one of the nucleic acids of SEQ ID Nos 1-4 and 7-8 and the sample to be assayed;

b) detecting the hybrid complex formed between the probe or the plurality of probes and the nucleic acid in the sample.

28. The method of claim 27, wherein said nucleic acid probe or the plurality of nucleic acid probes is selected from the group consisting of SEQ ID Nos 9-10, 11-12 and 13-14.

29. The method of claim 27, wherein said nucleic acid probe or the plurality of nucleic acid probes has been imobilized on a substrate.

30. The method of claim 27, wherein said nucleic acid probe or the plurality of nucleic acid probes is labeled with a detectable molecule.

31. A kit for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 1-4 and 7-8 in a sample, said kit comprising :

a) a nucleic acid probe or a plurality of nucleic acid probes which can hybridize to a nucleotide sequence included in one of the nucleic acids of SEQ ID Nos 1-4 and 7-8;

b) optionally, the reagents necessary for performing the hybridization reaction

32. The kit of claim 31, wherein said nucleic acid probe or said plurality of nucleic acid probes is selected from the group consisting of SEQ ID Nos 9-10, 11-12 and 13-14.

33. The kit of claim 31, wherein said nucleic acid probe or said plurality of nucleic acid probes is labeled with a detectable marker.

34. The kit of claim 31, wherein said nucleic acid probe or said plurality of nucleic acid probes is immobilized on a substrate.

35. An array of nucleic acid sequences comprising at least one of the sequences selected from the group consisting of SEQ ID Nos 7-8, 9-10, 11-12 and 13-14 or the sequences complementary thereto or a fragment thereof of at least 8 consecutive nucleotides thereof.

36. An array of nucleic acid sequences comprising at least two of the sequences selected from the group consisting of SEQ ID Nos 7-8, 9-10, 11-12 and 13-14 or the sequences complementary thereto or a fragment thereof of at least 8 consecutive nucleotides thereof.

37. A method for the detection in an individual of alleles of *TBC-1* associated with prostate cancer, wherein said method comprises the following steps :

-a) obtaining a nucleic acid sample from the individual to be tested, and

-b) determining the presence in the sample of an allele of a biallelic marker or of a group of biallelic markers of the *TBC-1* gene which, when taken alone or in combination with another/other biallelic marker/s of the *TBC-1* gene, is indicative of prostate cancer, of an early onset of prostate cancer, of the level of aggressiveness of prostate cancer tumors, of a modified or forthcoming expression of the *TBC-1* gene, of a modified or forthcoming production of the TBC-1 protein, or of the production of a modified TBC-1 protein.

38. The method of claim 37, wherein the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8, or a fragment thereof including the polymorphic base.

39. A diagnostic kit for determining the presence in a DNA sample of alleles associated with prostate cancer, with an early onset of prostate cancer, with the level of aggressiveness of prostate cancer tumors, with a modified expression of the *TBC-1* gene, with a modified production of the TBC-1 protein, or with the production of a modified TBC-1 protein, wherein said kit comprises primers for the amplification of the *TBC-1* gene.

40. A kit according to claim 39, wherein at least one of the primers is complementary to a nucleotide sequence of the *TBC-1* gene comprising an allele of a biallelic marker associated with prostate cancer.

41. A kit according to claim 39, wherein the primers comprise a nucleic acid selected from the group consisting of the nucleotide sequences of SEQ ID Nos 9-14.

42. A method for the treatment of prostate cancer comprising the following steps :
- selecting an individual whose DNA comprises an allele of a biallelic marker or of a group of biallelic markers, preferably markers of the *TBC-1* gene, associated with a susceptibility to prostate cancer;
- monitoring in said individual the appearance and optionally the development of a tumor in the prostate; and
- administering an effective amount of a medicament acting against prostate cancer to said individual at an appropriate stage of the prostate cancer.

43. The method of claim 42, wherein the biallelic marker comprises at least one of the sequences of SEQ ID Nos 7-8.

44. A recombinant vector comprising a nucleic acid encoding the TBC-1 polypeptide of the amino acid sequence of SEQ ID No 5, or a peptide fragment or variant thereof.

45. A recombinant vector comprising a nucleic acid according any one of claims 1 to 11.

46. A recombinant expression vector comprising a nucleic acid selected from the group consisting of SEQ ID No 2, or a biologically active fragment or variant thereof.

47. A recombinant expression vector comprising :
a) a nucleic acid comprising a regulatory polynucleotide of the nucleotide sequence of SEQ ID No 2, or a biologically active fragment or variant thereof;
b) a polynucleotide encoding a polypeptide or a polynucleotide of interest which is operably linked with said nucleic acid;
c) optionally, a nucleic acid comprising a 3'-regulatory polynucleotide, preferably a 3'-regulatory polynucleotide of the *TBC-1* gene, or a biologically active fragment or variant thereof.

48. A recombinant vector useful for the expression of the *TBC-1* coding sequence, wherein said vector comprises a nucleic acid selected from the group of SEQ ID Nos 1, 3 and 4 or a nucleic acid having at least 95% nucleotide identity with a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1, 3 and 4.

49. A recombinant expression vector comprising a nucleic acid comprising the nucleotide sequence beginning at the nucleotide in position 176 and ending in position 3733 of the polynucleotide of SEQ ID No 4.

50. A recombinant host cell comprising a nucleic acid coding for the TBC-1 polypeptide of SEQ ID No 5.

51. A recombinant host cell comprising a purified or isolated nucleic acid encoding a TBC-1 polypeptide, or a polypeptide fragment or variant thereof.

52. A recombinant host cell comprising a purified or isolated nucleic comprising at least 20 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4.

53. A recombinant host cell comprising a purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID No 2 or a biologically active fragment or variant of the nuclectide sequence of SEQ ID No 2.

54. A recombinant host cell comprising a purified or isolated nucleic acid comprising a 3'-regulatory sequence of the *TBC-1* gene, or a biologically active fragment or variant thereof.

55. A recombinant host cell comprising a polynucleotide consisting of :

(1) a nucleic acid comprising a regulatory polynucleotide of SEQ ID No 2 or a biologically active fragment or variant thereof;

(2) a polynucleotide encoding a desired polypeptide or nucleic acid.

(3) Optionally, a nucleic acid comprising a 3'-regulatory sequence, preferably a 3'-regulatory sequence of the *TBC-1* gene, or a biologically active fragment or variant thereof, wherein sequences (1), (2) and (3) are operably linked one with each other.

56. A recombinant host cell comprising a recombinant vector according to any one of claims 44 to 49.

57. A transgenic animal comprising in its somatic cells and/or its germ line cells a polynucleotide selected from the following group of polynucleotides :

a) a purified or isolated nucleic acid encoding a TBC-1 polypeptide, or a polypeptide fragment or variant thereof;

b) a purified or isolated nucleic comprising at least 20 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 3 and 4;

c) a purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID No 2 or a biologically active fragment or variant of the nucleotide sequence of SEQ ID No 2;

d) a purified or isolated nucleic acid comprising a 3'-regulatory sequence, preferably a 3'-regulatory sequence of the *TBC-1* gene, or a biologically active fragment or variant thereof;

e) a polynucleotide consisting of :

(1) a nucleic acid comprising a regulatory polynucleotide of SEQ ID No 2 or a biologically active fragment or variant thereof;

(2) a polynucleotide encoding a desired polypeptide or nucleic acid;

(3) Optionally, a nucleic acid comprising a 3'-regulatory sequence, preferably a 3'-regulatory sequence of the *TBC-1* gene, or a biologically active fragment or variant thereof, wherein sequences (1), (2) and (3) are operably linked one with each other.

58. A method for producing a polypeptide comprising the amino acid sequence of SEQ ID No 5 or a fragment or a variant thereof, wherein said method comprises the steps of :

a) culturing, in an appropriate culture medium, a cell host previously transformed or transfected with the recombinant vector of step b);

b) harvesting the culture medium thus conditioned or lyse the cell host, for example by sonication or by an osmotic shock;

c) separating or purifying, from the said culture medium, or from the pellet of the resultant host cell lysate the thus produced polypeptide of interest.

d) Optionally characterizing the produced polypeptide of interest

59. A TBC-1 polypeptide comprising an amino acid sequence of SEQ ID No 5, or a fragment or variant thereof.

60. The polypeptide of claim 59, wherein said polypeptide comprises the TBC domain of TBC-1.

61. The polypeptide of claim 59, wherein said polypeptide comprises the amino acid sequence EVGYCQGL.

62. A polypeptide comprising amino acid changes ranging from 1, 2, 3, 4, 5, 10 to 20 substitutions, additions or deletions of one amino acid as regards to the TBC-1 polypeptide of the amino acid sequence of SEQ ID No 5.

63. A polyclonal or a monoclonal antibody, a humanized antibody or a single chain Fv thereof, directed against the TBC-1 polypeptide of the amino acid sequence of SEQ ID No 5, or to a peptide fragment or variant thereof.

64. A method for detecting specifically the presence of a TBC-1 polypeptide in a biological sample, said method comprising the following steps :

a) bringing into contact the biological sample with an antibody directed against the TBC-1 polypeptide of the amino acid sequence of SEQ ID No 5, or to a peptide fragment or variant thereof;

b) detecting the antigen-antibody complex formed.

65. A diagnostic kit for detecting *in vitro* the presence of a TBC-1 polypeptide in a biological sample, wherein said kit comprises :

a) a polyclonal or monoclonal antibody directed against the TBC-1 polypeptide of the amino acid sequence of SEQ ID No 5, or to a peptide fragment or variant thereof, optionally labeled;

b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent carrying optionally a label, or being able to be recognized itself by a labeled reagent, more particularly in the case when the monoclonal or polyclonal antibody is not itself labeled.

66. A method for the screening of candidate substances for prostate cancer treatment, wherein said method comprises the following steps:

a) providing a cell line, an organ, or a mammal expressing a *TBC-1* gene or a fragment thereof, preferably the regulatory region or the promoter region of the *TBC-1* gene.

b) obtaining a candidate substance;

c) testing the ability of the candidate substance to decrease the symptoms of prostate cancer and/or to modulate the expression levels of *TBC-1*.

67. A method for the screening of a candidate substance interacting with the TBC-1 polypeptide, wherein said method comprises the following steps :

a) providing a polypeptide comprising the amino acid sequence SEQ ID No 5, or a peptide fragment or a variant thereof;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance;

d) detecting the complexes formed between said polypeptide and said candidate substance.

68. A kit for the screening of a candidate substance interacting with the TBC-1 polypeptide, wherein said kit comprises :

a) a TBC-1 protein having the amino acid sequence of SEQ ID No 5 or a peptide fragment or a variant thereof ;

b) optionally means useful to detect the complex formed between the TBC-1 protein or its peptide fragment or variant and the candidate substance.

69. The kit of claim 68 wherein the detection means consist in monoclonal or polyclonal antibodies directed against the TBC-1 protein or a peptide fragment or a variant thereof.

70. A method for the screening of a candidate substance or molecule that modulates the expression of the *TBC-1* gene, wherein this method comprises the following steps :

a) providing a recombinant host cell containing a nucleic acid, wherein said nucleic acid comprises the nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof, the nucleotide sequence of SEQ ID No 2 or its biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein;

b) obtaining a candidate substance, and

c) determining the ability of the candidate substance to modulate the expression levels of the polynucleotide encoding the detectable protein.

71. The screening method of claim 70, wherein the nucleic acid comprising the nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof also includes a 5'UTR region of one of the *TBC-1* cDNAs of SEQ ID Nos 3 and 4, or one of their biologically active fragments or variants thereof.

72. A kit useful for the screening of a candidate substance or molecule that modulates the expression of the *TBC-1* gene, wherein said kit comprises :

a) a recombinant vector that allows the expression of a nucleic acid comprising a nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof;

b) a polynucleotide encoding a detectable protein which is operably linked to the nucleotide sequence of SEQ ID No 2 or a biologically active fragment or variant thereof.

73. A method for the screening of a candidate substance or molecule that modulates the expression of the *TBC-1* gene, wherein siad method comprises the following steps :

a) providing a recombinant host cell containing a nucleic acid, wherein said nucleic acid comprises a 5'UTR sequence of one of the *TBC-1* cDNAs of SEQ ID Nos 3 and 4, or one of their biologically active fragments or variants, the 5'UTR sequence or its biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein;

b) obtaining a candidate substance, and;

c) determining the ability of the candidate substance to modulate the expression levels of the polynucleotide encoding the detectable protein.

74. The method of claim 73, wherein the nucleic acid that comprises a nucleotide sequence selected from the group consisting of the 5'UTR sequence of one of the *TBC-1* cDNAs of SEQ ID Nos 3 and 4 or one of their biologically active fragments or variants, includes a promoter sequence which is endogenous with respect to the *TBC-1* 5'UTR sequences.

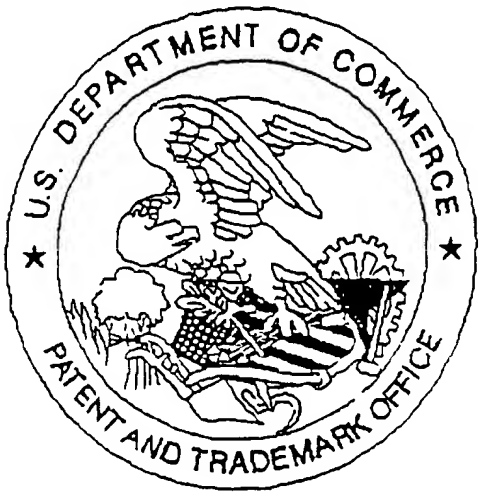
75. The method of claim 73, wherein the nucleic acid that comprises a nucleotide sequence selected from the group consisting of the 5'UTR sequence of one of the *TBC-1* cDNAs of SEQ ID Nos 3 and 4 or one of their biologically active fragments or variants, includes a promoter sequence which is exogenous with respect to the *TBC-1* 5'UTR sequences defined therein.

76. A kit for the screening of a candidate substance modulating the expression of the *TBC-1* gene, wherein said kit comprises :

a) a recombinant vector that comprises a nucleic acid including a 5'UTR sequence of one of the *TBC-1* cDNAs of SEQ ID Nos 3 and 4, or one of their biologically active fragments or variants, the 5'UTR sequence or its biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein.

United States Patent & Trademark Office

Office of Initial Patent Examination -- Scanning Division



Application deficiencies found during scanning:

1. Application papers are not suitable for scanning and are not in compliance with 37 CFR 1.52 because:
   - ☐ All sheets must be the same size and either A4 (21 cm x 29.7 cm) or 8-1/2"x 11" Pages _____ do not meet these requirements.
   - ☐ Papers are not flexible, strong, smooth, non-shiny, durable, and white.
   - ☐ Papers are not typewritten or mechanically printed in permanent ink on one side.
   - ☐ Papers contain improper margins. Each sheet must have a left margin of at least 2.5 cm (1") and top, bottom and right margins of at least 2.0 cm (3/4").
   - ☐ Papers contain hand lettering.

2. Drawings are not in compliance and were not scanned because:
   - ☐ The drawings or copy of drawings are not suitable for electronic reproduction.
   - ☐ All drawings sheets are not the same size. Pages must be either A4 (21 cm x 29.7 cm) or 8-1/2" x 11".
   - ☐ Each sheet must include a top and left margin of at least 2.5 cm (1"), a right margin of at least 1.5 cm (9/16") and a bottom margin of at least 1.0 cm (3/8").

3. Page(s)_____ are not of sufficient clarity, contrast and quality for electronic reproduction.

4. Page(s)_____ are missing.

5. OTHER: No Declarations

Human TBC-1 partial genome structure and cDNAs

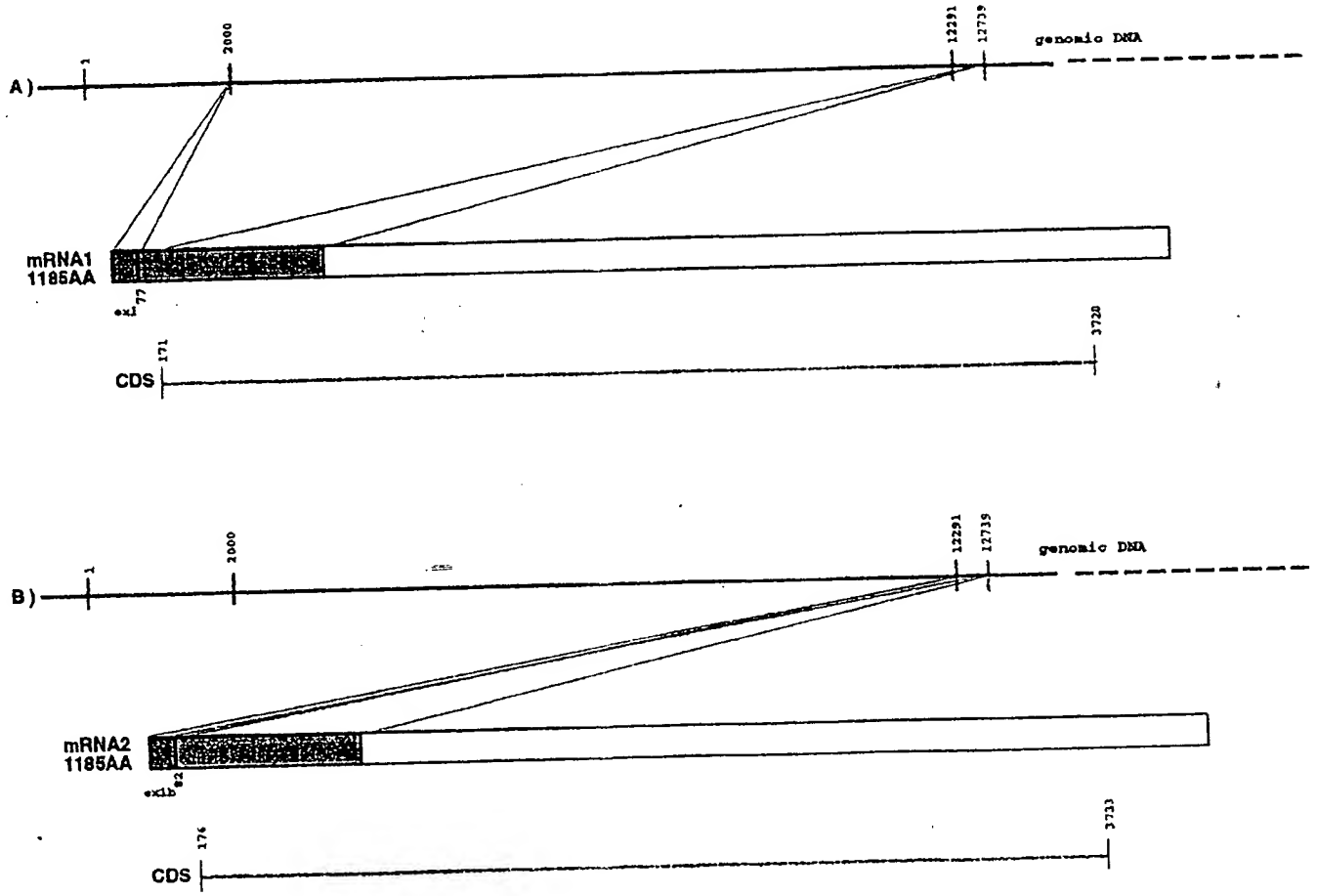
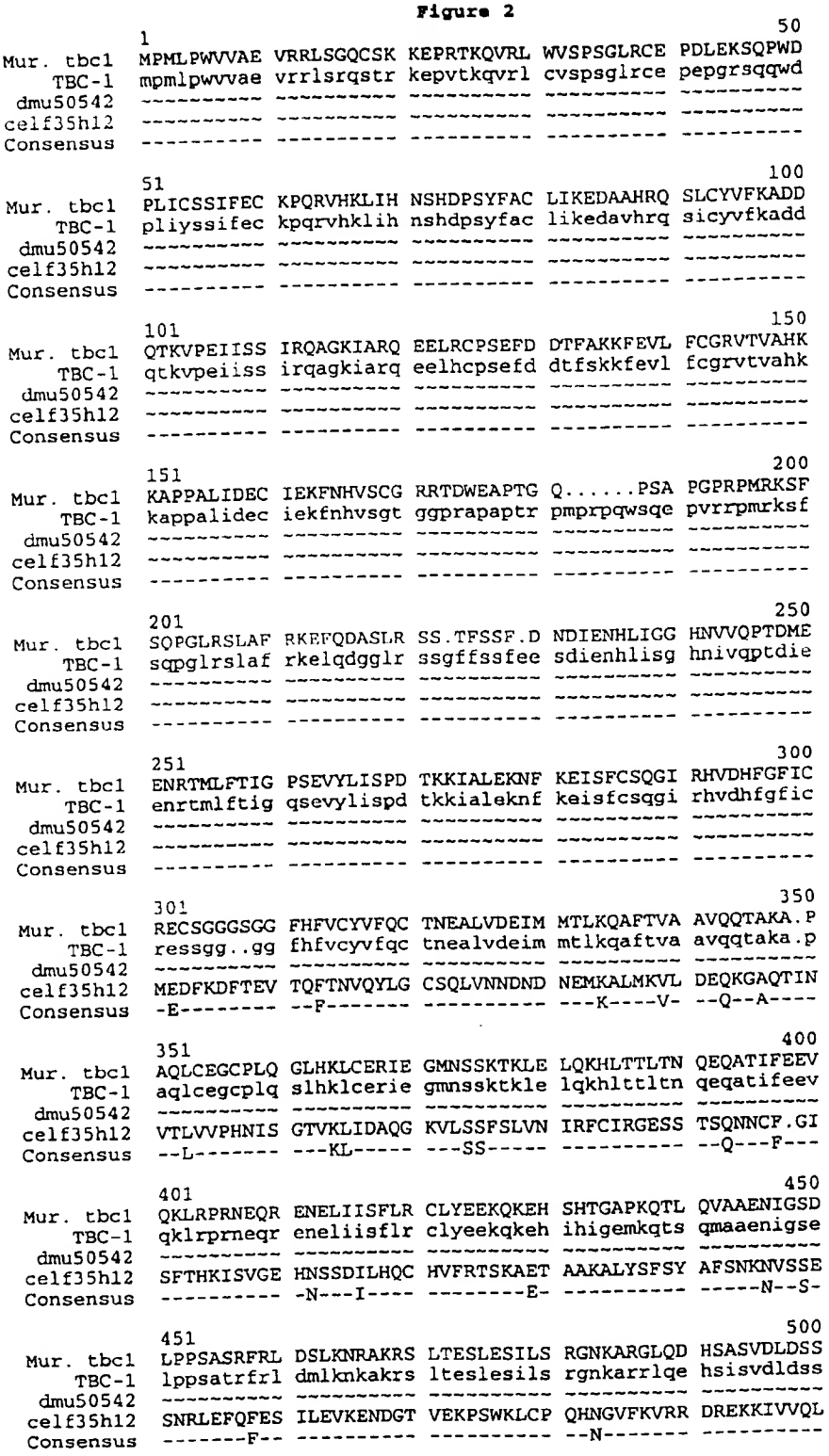


FIGURE 1

Figure 2



```
                                                                      50
                1 MPMLPWVVAE VRRLSGQCSK KEPRTKQVRL WVSPSGLRCE PDLEKSQPWD
Mur. tbc1       1 MPMLPWVVAE VRRLSGQCSK KEPRTKQVRL WVSPSGLRCE PDLEKSQPWD
    TBC-1       1 mpmlpwvvae vrrlsrqstr kepvtkqvrl cvspsglrce pepgrsqqwd
 dmu50542       1 ---------- ---------- ---------- ---------- ----------
celf35h12       1 ---------- ---------- ---------- ---------- ----------
Consensus       1 ---------- ---------- ---------- ---------- ----------

                                                                     100
               51 PLICSSIFEC KPQRVHKLIH NSHDPSYFAC LIKEDAAHRQ SLCYVFKADD
Mur. tbc1      51 PLICSSIFEC KPQRVHKLIH NSHDPSYFAC LIKEDAAHRQ SLCYVFKADD
    TBC-1      51 pliyssifec kpqrvhklih nshdpsyfac likedavhrq sicyvfkadd
 dmu50542      51 ---------- ---------- ---------- ---------- ----------
celf35h12      51 ---------- ---------- ---------- ---------- ----------
Consensus      51 ---------- ---------- ---------- ---------- ----------

                                                                     150
              101 QTKVPEIISS IRQAGKIARQ EELRCPSEFD DTFAKKFEVL FCGRVTVAHK
Mur. tbc1     101 QTKVPEIISS IRQAGKIARQ EELRCPSEFD DTFAKKFEVL FCGRVTVAHK
    TBC-1     101 qtkvpeiiss irqagkiarq eelhcpsefd dtfskkfevl fcgrvtvahk
 dmu50542     101 ---------- ---------- ---------- ---------- ----------
celf35h12     101 ---------- ---------- ---------- ---------- ----------
Consensus     101 ---------- ---------- ---------- ---------- ----------

                                                                     200
              151 KAPPALIDEC IEKFNHVSCG RRTDWEAPTG Q......PSA PGPRPMRKSF
Mur. tbc1     151 KAPPALIDEC IEKFNHVSCG RRTDWEAPTG Q......PSA PGPRPMRKSF
    TBC-1     151 kappalidec iekfnhvsgt ggprapaptr pmprpqwsqe pvrrpmrksf
 dmu50542     151 ---------- ---------- ---------- ---------- ----------
celf35h12     151 ---------- ---------- ---------- ---------- ----------
Consensus     151 ---------- ---------- ---------- ---------- ----------

                                                                     250
              201 SQPGLRSLAF RKEFQDASLR SS.TFSSF.D NDIENHLIGG HNVVQPTDME
Mur. tbc1     201 SQPGLRSLAF RKEFQDASLR SS.TFSSF.D NDIENHLIGG HNVVQPTDME
    TBC-1     201 sqpglrslaf rkelqdgglr ssgffssfee sdienhlisg hnivqptdie
 dmu50542     201 ---------- ---------- ---------- ---------- ----------
celf35h12     201 ---------- ---------- ---------- ---------- ----------
Consensus     201 ---------- ---------- ---------- ---------- ----------

                                                                     300
              251 ENRTMLFTIG PSEVYLISPD TKKIALEKNF KEISFCSQGI RHVDHFGFIC
Mur. tbc1     251 ENRTMLFTIG PSEVYLISPD TKKIALEKNF KEISFCSQGI RHVDHFGFIC
    TBC-1     251 enrtmlftig qsevylispd tkkialeknf keisfcsqgi rhvdhfgfic
 dmu50542     251 ---------- ---------- ---------- ---------- ----------
celf35h12     251 ---------- ---------- ---------- ---------- ----------
Consensus     251 ---------- ---------- ---------- ---------- ----------

                                                                     350
              301 RECSGGGSGG FHFVCYVFQC TNEALVDEIM MTLKQAFTVA AVQQTAKA.P
Mur. tbc1     301 RECSGGGSGG FHFVCYVFQC TNEALVDEIM MTLKQAFTVA AVQQTAKA.P
    TBC-1     301 ressgg..gg fhfvcyvfqc tnealvdeim mtlkqaftva avqqtaka.p
 dmu50542     301 ---------- ---------- ---------- ---------- ----------
celf35h12     301 MEDFKDFTEV TQFTNVQYLG CSQLVNNDND NEMKALMKVL DEQKGAQTIN
Consensus     301 -E-------- --F------- ---------- ---K----V- --Q--A----

                                                                     400
              351 AQLCEGCPLQ GLHKLCERIE GMNSSKTKLE LQKHLTTLTN QEQATIFEEV
Mur. tbc1     351 AQLCEGCPLQ GLHKLCERIE GMNSSKTKLE LQKHLTTLTN QEQATIFEEV
    TBC-1     351 aqlcegcplq slhklcerie gmnssktkle lqkhlttltn qeqatifeev
 dmu50542     351 ---------- ---------- ---------- ---------- ----------
celf35h12     351 VTLVVPHNIS GTVKLIDAQG KVLSSFSLVN IRFCIRGESS TSQNNCF.GI
Consensus     351 --L------- ---KL----- ---SS----- ---------- --Q---F---

                                                                     450
              401 QKLRPRNEQR ENELIISFLR CLYEEKQKEH SHTGAPKQTL QVAAENIGSD
Mur. tbc1     401 QKLRPRNEQR ENELIISFLR CLYEEKQKEH SHTGAPKQTL QVAAENIGSD
    TBC-1     401 qklrprneqr eneliisflr clyeekqkeh ihigemkqts qmaaenigse
 dmu50542     401 ---------- ---------- ---------- ---------- ----------
celf35h12     401 SFTHKISVGE HNSSDILHQC HVFRTSKAET AAKALYSFSY AFSNKNVSSE
Consensus     401 ---------- -N---I---- --------E- ---------- -----N--S-

                                                                     500
              451 LPPSASRFRL DSLKNRAKRS LTESLESILS RGNKARGLQD HSASVDLDSS
Mur. tbc1     451 LPPSASRFRL DSLKNRAKRS LTESLESILS RGNKARGLQD HSASVDLDSS
    TBC-1     451 lppsatrfrl dmlknkakrs lteslesils rgnkarrlqe hsisvdldss
 dmu50542     451 ---------- ---------- ---------- ---------- ----------
celf35h12     451 SNRLEFQFES ILEVKENDGT VEKPSWKLCP QHNGVFKVRR DREKKIVVQL
Consensus     451 -------F-- ---------- ---------- --N------- ----------
```

Figure 2 (Continued I)

```
                                                                      550
           501
Mur. tbc1  TSSTLSNTSK ELSMGDKEAF PVSETSFKLL GSSDDLSSDS EGHIAEESAL
    TBC-1  lsstlsntsk epsvcekeaf pisessfkll gssedlssds eshlpeepap
 dmu50542  ---------- ---------- ---------- ---------- ----------
celf35h12  RQVARKKTID GFLLNIKKCF GMLLAAGRNL RHSDLQLLEM DRNATGTDSA
Consensus  -------T-- ------K--F ---------L --S------- ----------

                                                                      600
           551
Mur. tbc1  LSPQQAFRRR ANTLSHFPVE CPAPPEPAQS SPGVSQRKLM RYHSVSTETP
    TBC-1  lspqqafrrr antlshfpie cqeppqparg spgvsqrklm ryhsvstetp
 dmu50542  ---------- ---------- ---------- ---------- ----------
celf35h12  VFVIEA..NW DPRVHMFEVL NTETPRDTRV FMTVAIDVIV SEISEPIRFS
Consensus  -----A---- ------F--- ----P----- ---V------ ---S------

                                                                      650
           601
Mur. tbc1  HERKDFESKA NHLGDTDGTP VKTRRHSWRQ QIFLRVATPQ KACDSPSRYE
    TBC-1  herkdfeska nhlgdsggtp vktrshswrq qiflrvatpq kacdsssrye
 dmu50542  ---------- ---------- ---------- ---------- ----------
celf35h12  MEAMSRVFHE HERFYKTPQT VVSEEFTLVL EVRIRTEKLL ETHGNMLLKS
Consensus-E-------- ---------- V--------- ----R----- ----------

                                                                      700
           651
Mur. tbc1  DYSELGELPP RSPLEPVCED GPFGQYRKKR GRRHASFESC GKRPSCSRSC
    TBC-1  dyselgelpp rsplepvced gpfaphqrkr kghlvssesc akrlffnryc
 dmu50542  ---------- ---------- --MRKPAKRG KRDAAELREL WRTAIRQTIM
celf35h12  VPIDFAWQLE GVYFLPTPSK SCDQSDPNDR KLTFISLESD SDRKRSKQNL
Consensus  ---------- -----P---- ---------R K----S-ES- --R-------

                                                                      750
           701
Mur. tbc1  ...LVRMEKE NQKLQASEND LLNKRLKLDY .EEITPCLKE VTTVWEKMLS
    TBC-1  ...clgmeke nqklqasend llnkrlklrf mkeitpclke vttvwekmls
 dmu50542  ...LNRMETE NAMLQARQNE NELKRIKLDY .EEIVPCDKQ LIERWEQIIE
celf35h12  GKSPSRMPTQ LLHPTGDDES DCDEPLLSGS GKVSQECKEE HLEMWDQLIE
Consensus  -----RME-E N--LQA--N- ---KRLKL-- --EI-PC-KE ----WE----

           751                                                        800
Mur. tbc1  .TPGRSK
    TBC-1  .tpgrsk
 dmu50542  RNSTQIG
celf35h12  NWDQQSD
Consensus  -----S-

           801                                                        850
Mur. tbc1
    TBC-1
 dmu50542
celf35h12
Consensus

           851                                                        900
Mur. tbc1  EVGYCQGL
    TBC-1  evgycqgl
 dmu50542  ELGYCQGL
celf35h12  EVSYCQGL
Consensus  EVGYCQGL

           901                                                        950
Mur. tbc1                                                          A
    TBC-1                                                          a
 dmu50542                                                          S
celf35h12                                                          T
Consensus                                                          -

                                                                     1000
           951
Mur. tbc1  SQFPLGFVAR VFDMIFLQGS EVIFKVALSL LGSHKPLILQ HENLETIVDF
    TBC-1  sqfplgfvar vfdmiflqgt evifkvavsl lgshkplilq henletivdf
 dmu50542  SQFPLGFVAR VFDLLFLESS DVIFKFAIAL LSVHKQQLLA KDNFEEIMDY
celf35h12  AKFPLQMVFF ILDLFLSQGM NTIFHISLAL LDDAKTDLLQ LD.FEGTLKY
Consensus  SQFPLGFVAR VFD--FLQG- -VIFK-A--L L--HK---LQ --N-E-I-D-
```

Figure 2 (Continued II)

```
                1001                                                1050
Mur. tbc1       IKNTLPNLGL VQMEKTISQV FEMDIAKQLQ AYEVEYHVVQ EELIESSPLS
    TBC-1       ikstlpnlgl vqmektinqv femdiakqlq ayevehhvlq eelidsspls
 dmu50542       LKTVVPKMEH TCMEQIMKLV FSMDIGKQLA EYNVEYNVLQ EEI.....TT
celf35h12       FRVSLPRKYR T..EASTKCL IHKAVKFRLN HSKLEVYENE YKRIKELERE
Consensus       -K--LP---- --ME-----V F-MDI-KQL- -Y-VE--V-Q EE-I------

                1051                                                1100
Mur. tbc1       DNQRMEKLEK TNSTLRKQNL DLL.EQLQVA NARIQSLEAT VEKLLTSESK
    TBC-1       dnqrmdklek tnsslrkqnl dll.eqlqva ngriqsleat iekllssesk
 dmu50542       TNHHLEMLNR E....KTQNQ HLE.QQLQFA QSSIAQLETT RSSQQAQITT
celf35h12       NEDPVLRMEK EIGRHQANTL RLERENDDLA HELVTSKIEL RRKLDVAEDQ
Consensus       -N-----LEK -------QNL -L--EQLQ-A ---I-SLE-T --KL---E--

                1101                                                1150
Mur. tbc1       LKQRALTLEV ERRPAADGGG AAEAKRPAQH SR.ARLHPAG AHRRLTAAR.
    TBC-1       lkqamltlel ersallqtve elrrrsaeps drepectqpr ahgrltalqe
 dmu50542       LQSQVQSLEL TIQTLGRYVG QLVEHNP... DLELPNEVRR MLQQLDDLDR
celf35h12       IETSANAIER LTRQNMDILE E..NKNLMRE YEQIKEMYRR DVLRLEENGS
Consensus       L  ------LE- ---------- ---------- ---------R ---RL-----

                1151                                                1200
Mur. tbc1       RDCAPLTLSK P~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    TBC-1       rlqhhptlsr p~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 dmu50542       QRRKPIFTER KIGKSVSVNS HLGFPLKVLE ELTERDELGS PQKQKKEKTP
celf35h12       RAEKLLAEYK KLFSERSKRA ENEREHFEVQ KKAIIARISD CDKCWPAVCE
Consensus       R--------- ---------- ---------- ---------- ----------

                1201                                                1250
Mur. tbc1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    TBC-1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 dmu50542       FFEQLRQQQQ QHRLNGGGQS SNVGESGSPT PPSRPNRLLD NASARTVMQV
celf35h12       .WEKNRSPVH SASTPTGPDL LTKLEEREDH IKNLEIDLAQ TKLSLVEAEC
Consensus       ---------- ---------- ---------- ---------- ----------

                1251                                                1300
Mur. tbc1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    TBC-1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 dmu50542       KLDELKLPEH VDKFVANIKS PLEVDSGVGT PLSPPSTASN SSGGSIFSRM
celf35h12       RNQDLTHQLM AQSESDGKKW FKKTITQLKE VGSSLKHHER SNSSVTPHFS
Consensus       ---------- ---------- ---------- ---------- ----------

                1301                                                1350
Mur. tbc1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    TBC-1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 dmu50542       GYRTTPPALS PLAQRQSYGV AITTAPCPQH MEEVAPATTM AVMPQEDVEE
celf35h12       STFQLQMDHT ETTTSNNIGY NSSSESFAVR FMQTPSAVLK ITNGEMTEDN
Consensus       ---------- ---------- ---------- ---------- ----------

                1351                                                1400
Mur. tbc1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    TBC-1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 dmu50542       PQPMHPLSMV GGDVNVRFKG TTQLKSIRPV HHMRAIPLGG VQHPSSTEPA
celf35h12       NNMLHGINGV DLLDLQSTDN DDQYSNSSSL ESRNSLTNHQ GKAEDSTMVT
Consensus       ---------- ---------- ---------- ---------- ----------

                1401                                                1450
Mur. tbc1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    TBC-1       ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 dmu50542       VRVAPVPVEL APPAATATTG RS~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
celf35h12       VNLDQLPARR TLLKCLVLVV FGKSYTKVIF QMLRFSGFLM RRGVYPGTLY
Consensus       ---------- ---------- ---------- ---------- ----------

                1451       1465
Mur. tbc1       ~~~~~~~~~~ ~~~~~
    TBC-1       ~~~~~~~~~~ ~~~~~
 dmu50542       ~~~~~~~~~~ ~~~~~
celf35h12       FRRCSKILLK KYDRI
Consensus       ---------- -----
```

(1) GENERAL INFORMATION:

(i) APPLICANT: Bougueleret, Lydie
Blumenfeld, Marta
Chumakov, Ilya

(ii) TITLE OF INVENTION: nucleic acids encoding human TBC-1 protein and polymorphic markers thereof.

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDANCE ADDRESS:
(A) ADDRESSEE: Knobbe, Martens, Olson & Bear
(B) STREET: 501 West Broadway
(C) CITY: San Diego
(D) STATE OR PROVINCE: California
(E) COUNTRY: USA
(F) ZIP: 92101-3505

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy Disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: Win95
(D) SOFTWARE: Word

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Israelsen, Ned A.
(B) REGISTRATION NUMBER: 29,655
(C) REFERENCE/DOCKET NUMBER: GENSET.037PR

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (619) 235-8550
(B) TELEFAX: (619) 235-0176

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17589 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: Polymorphic fragment 99-430-352 SEQ ID7
(B) LOCATION: complement 9470..9516

(ix) FEATURE:
(A) NAME/KEY: Polymorphic fragment 99-430-352 SEQ ID8
(B) LOCATION: complement 9470..9516

(ix) FEATURE:
(A) NAME/KEY: exon1
(B) LOCATION: 2000..2076

(ix) FEATURE:
(A) NAME/KEY: exon1b
(B) LOCATION: 12291..12372

(ix) FEATURE:
(A) NAME/KEY: exon2
(B) LOCATION: 12739..13248

(ix) FEATURE:
(A) NAME/KEY: homology with EST in :SEQID6
(B) LOCATION: 12291..12372

(ix) FEATURE:
(A) NAME/KEY: homology with EST in :SEQID6
(B) LOCATION: 12739..13114

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGACAGTATC TAGCACAATA CCCCAAATCG ACTAACTCCT CCGTAAAGAA TAGCTACCAC 60

TATTGTGAGA GTTTTAAGTC AAGCTGTGAA TAAAACTCTT GGGTCCACTT AAAAATACCT 120

CCCCTGGATG TAAGCATCCA GGGAAATCAG GGAATGCCAT AAGACAGCCC TAATCTAAAA 180

GCCTACAAGA AGCTCAGTGG GCTTCAAGGA AGACACTGCT CTTGGTACGA TGAGGAAACC 240

TGGCCCTCTA TTTGCCTCCT GGGCCACAGT AATATTGATA ATAGCTGCTG CTTTTAGTTG 300

AGGACCATGT ACGTCTGTGT CACTGCACTG GCCACTTTAC TTACACTTTC CTGCTTTGTC 360

CTCACAAAGA TCCTGTAAGG TGTGTATTGG TCCCATTTAG CAGGTAAGAC AATGAAGACC 420

AGAGGTCCAG CACCTTGCCT AAACCACACC TGCTGGGATT TGGATTCAAG TCCAACCGTA 480

CAGCTCAAAC GCTCAGCCAC TTCCCTAAAG TCCACCCCCA GCTACATTAA GTAAAAAAAT 540

CCAGAAAGAT GCCACCTGGG GGTCTGGAAC TGCCTCCTCC GAGCACCCGG CTCTCCCCTC 600

CCTGCGGACT CTTCTCTGGA GAGGATGTGA TGCTTCTTAC TTTTCTCAGA TCCCTCTCCC 660

CACCCTGCGA GTGACGTTGC GCCTCTGTGC CTGGTGGGAT AGGGATCTGG GAGCTTCGCC 720

TGTTTTTTGC ACACTGCCAT CCCCTAGTCT TAGGGAGCGA GCTCTGTCCC GCTTTTCACA 780

TCTCCGCGTC TTTCCTTGCA CTCTACATCA CCGCTGGGAA TGTCCCCAGA CCTGATCGGG 840

GCATGCACAC TGGGGTGTGC GTGTGCGTGT GGTGTGTGTT CCTGCGCGTG TGCCGGGCTC 900

GCGGGGCAGG AAAAAGCGCC TAATCCAGGC TCTGCGTCAC TCCCGCAATT GGTTAGAAAT 960

GGAGTTTCCT GGTGTTTAAT CCCGGGAGGG CACTTCGCCT TCGTTGTTTC CCAGAGTCCC 1020

TGATTTTCCT GCCTCGCATG CCAGCGCCCC ATAGGGCATC CGTGCCTCAG TTCACCTCTT 1080

GCCATCCTCC AAGGACGGGG AGAAGGGGTA AGGCGGGGGA GAGCAAGGTG GCTTGGTCGC 1140

CCCCGGCCCC CGCCCCCCAT GTTGTGTGCA GTTTCCACCA CGTCTGTTTC GGAGGGAGAA 1200

GAGGAGGGTG CAGATGAGGC GAGGCGCCTT CGGGAGCGCG GAGAGCGGGC AGGCAGTGCC 1260

ACCTGCTGAG AGCCACTCAG GCCGAGCAAG CGGCGGGCAG TGCCACCTGC TATAAATAGG 1320

CCGCCAAGGA CAGGGTGTGC GACTGTACAT CCCGCCACGA GGGCCTGCAT CACGCGCGGG 1380

GCCCCGCGCC CCCGGCTCCC CAGGGAAACG CTGTGCCCAG ATCCTGCGCA GGGGTCTGGA 1440

TGGGGCGGCG GCCCGAGTAC TTCCCCCCTA TTCCCCCCAC ANACACTGGC TGAGGATGGC 1500

CCGCGGGCTT GGGGGCGGGG GGTGGCAAGG AGGGGAGGGA GGCCGCGGCG GACCCGCAGT 1560

GCAGCAGCTG TTGCTCGCGT GTGACTCGCC CGTCCGGGCC GTGCTGCCCA GGCACAGTCA 1620

CACGGCGCAG TGGGGAGGAG GAGGACACCG AGTCCCCCTC CCAGCTCCCC GGGGACCGAG 1680

TGGGGAGATC CCGGCTCCTG TCTTCCCCTC GCCTCCAGCG CGCTCGCCCA GGCTGGGAGG 1740

AGGAAACCAG AGCCGCGCGC AGACACCTCC TCCTTCTCCT CCTCTTCTTC CTCCTCCTCC 1800

TCCTCCTCCT CCTCCTCTTC GGCTGCTGCT CCTGGTGCCG CCACCGTCCG CCGGTGCCTG 1860

TTGCTGCCGC CGCCGCGGGA CCTGCTGTGT CCTCAGCTGG GTGGAGAAGA GGCGGGCGCC 1920

GAGCCGAGGG GAGCCCCCTC CCCGTCCCCC CGCGGCGGGA AGAGCGCAGC CAGCCGGGTG 1980

CGATGGACTC CCCGCCCGCC CAGGCCGTCC CCAGGATGCC CCCAAGCACC TGCGCGTCCC 2040

GGCCCGGCCC CGGGCTCTGA GCGCGCCGCG GCACAGGTAA GGCGCTTCCT GGGGCTTCGT 2100

CCTGGCCACC CTGCTGGCTC CTCTCGGGGC GTCGCGGCCG CCCCCTCCCG CAGCACGCCC 2160

CTGCCCCGCC TGGCCGCGGA GGGGAAGGCA TCTGGCCGCC CACGGACGCG AGGCCAGGGT 2220

CTCTCGGGGG AGGAAGTTCA TTGCCATCTC GTTGCCCCCC TTACCCCCCC ACCCCCGCCG 2280

CCCTTGGACG AAAGCGAAAC CTTAATGTTG CTAGCGACCC GAGAGCTCCG CCGGCTTCTC 2340

CCCCAACCCC CGCCAGCTCA CTGGTCCGCG CATCTCTCCC CTCCCCCCTC CCGCCAATTA 2400

TCCTAGCGTG TTTGCAAGGC GACCAGATTG GAAAGAGTGT GGTCAGAGTG ACCCCAAGCC 2460

ACGCTTTAAA AGTTCAGGGT ACTTTGCAGT AGTAACTTTG GCAGCTCCAC CAGTGCGCGC 2520

AACATTTCTT TCTATGGGTA CATCCTGTAC CAGTCATTTT GAAACCCTGC TTCATTGTTT 2580

CTAGCCGCTT CCTGATGGCT CTGTGATTAT GAGACCCCCC TCAAACTTCA CCAGGCATTA 2640

AGGTTTTGTT TTTGCTTTTT TTTCAGAGAG GTATCATTTC GTTTGAAATC CACCTAGATG 2700

TGGCTTTTCC TGTTTTGATT TTACTTAACA TAGCTTATTC TCTGGAAGTT GCTTTAAAAA 2760

GAAATTGAAA GTGATGGTTG TTCCTTCCAC CAAACAGTTT AATTTTCAGG GTGCCTCATA 2820

TTAATGGATA TGTTTTCCCT TCATAGATTT CTCATTGTTT CCCTTATGAT GGGATGATTT 2880

CATTTATTAA TAAAATCAGA CTTTGAAAGA GCATTTAAAA ATGACCTGGT TTAAATAGGT 2940

CACACCCAAG AAACTCAGCT ATCTGTACAA GTTCAAACTT CTAAACTTTT TCAATGAGCT 3000

AGGGGTGGTG GCACCCACCT GTAGTCCCAG CTACTTGGGA GGCTGAGGCA GGAGGATCAC 3060

TTGAGCCCAG GAGTTCGAGG CCATAGTGAG CTATGACTGT GCCACCTCAC TGGAGCCTGG 3120

GTGACAAAGT GAGATCCCAT CTCTTAAAAA AAAAGAGTTT AGGGGACATT TTCTGAAGTG 3180

AACACAAGTA GAGCATTCTA ACACTATTGA GTGCAAGGAG ACCTGGAAGG GACTAAGTGG 3240

TTCAAAGCAG GAAATAAAAT CATCAGGTGA TAATTAAAAT AATTTCTTTC CTGTGGATTT 3300

GTCCAGCCAT TTGCAAACCA GGAGAATAGG AAAAAAAATC ACTAGTGTAG TTATAAATTA 3360

TTACATTACG TTTTCAAAGG AAAATTTTGC AAATGCGTCT CCTTGTCATA GTCTATTGTT 3420

ATCTACCCCA CTGAGAGTGC TGGGGCTTCC CCTTTTCACC ACGACAGCAT TTCTGGTTGG 3480

GTGGCAGTCA TGCAGTGTTG ACCTGGTGTC CCATAAGGCA CAGTTTGTCA AAACACTAGT 3540

GGGTATTAGG AGGAAACGTG CAACTCTGAA GCAACAGAGC TTGCCCCTTC TTCCTCATTA 3600

TCCAGCTGGT GATAATCCCT GTCCCCCACT TCCCTAGAAG ACAGCTTTGA CCAGGAAGGC 3660

TGCAATGACA ATGAGATGTA CCCCTATGCA GAGCCAGATG TGGGCGGGTG GCTTTTTTGT 3720

GGTCCAGATC TTCTAGGATC TTCTAGGATG TAACCCTGGC AAGCAGTGGG GAGCCTGAAT 3780

CAAGCAGCAT GGCTGTTACC TCTTCTGTGT TCACAGCAGC ATCTTCAGTT GTCTTGGTGC 3840

CTGGAGCAGG CACCACAGCT GCCTGCTCTG TTGGCCACCA GCTTTCTAGA GTAGATGGTA 3900

GGGAGGAGAG CAAGGGGCTC AAGAGGATTC TGTCTTTGAA CATGCTTTTA ANTTTGATCT 3960

GACAGAATGG CAGCTCCCTG AAGTCCTTCC TACTCTCTCC ACAGCATTTC TCTGTAGGTC 4020

CCCAGTTTTT GCTCTTTTCA GATTCCCAGA GGACNTGAAA ATGTATCACG GCCCATTTGG 4080

GGACTTCCTG TATATGTGTG GGTGCCTCAG GATCATTTGT TTTGCCCTTT TCCAGTCTAC 4140

CGTGCTGCCC TTCTCAAGTT TAATGNACCA CGTTAGTTTC AATATTTTAT ATATTTCTCA 4200

GCAGTTTTCA TCTCTTGGTC ATTAAACTTG AGAAGTAAAA TCTGCTCATT AAAATGACTG 4260

AGTCCATGGC CAGGCATGGT GGCTCATGCC TGTAATCCCA GCACTTTGGG AGTCCAAGGC 4320

GGGTGGATCA CTTGAGGTCA GGAGTTCGAG ACCAGCCTGG CCAGCATGGC AAAACCCTGT 4380

CTCTACAAAA ATATAGATCT ACAAAAACTA GCCAGGCATG GTGGCATGTG CCTGTAGTCC 4440

CAGCTATTTG GGAGGCTGAG ACAGGAGAAT CGCTTGAAGC CAGGAGGCGG AGGTTGCAGT 4500

GAAACATGAT CGTGCCACTG AGTCCATTCA GCAGCAGAGT AGTGTTGGGG TTTGTATCCC 4560

TGTAGTGATG ACGAAGGNTT TAGGTTTTCA GTCAGAACTG TTACCTTACA ATTTCCTTCA 4620

CTGACTTTTC TTCCTTTCCA ACACCACATT CCAATAAAAA ATATCTTTAG ACCAGATTCT 4680

TCACGAAAGA CATGAAGGTT TTCATGCTTC AAGGTTTTTG ACTTTTTTTT TTTTTTTNAA 4740

AGGAGTCTTG CTGTGTCACC CAGGCTGGAG TGCAGTGGCG TGATCTCAGC TCACTGCAAC 4800

CTCCGCCTCC TGGGTTCAAG TGATTCTCCT GCCTCAGCCT CCCAAGTAGC TGGGACTACA 4860

GGCGTGCTCT ACCACGGCCG GCTAAATTTT GTGTTTTTAG TAGAGGCGAG GTTTCACCAT 4920

CTTGGCCAGG CTGGTCTTGA ACTCCCGACC TTGTGATCCA CCCGCCTTGG CCTCCCAAAG 4980

TGCTGGGATT ACAGGTGTGA GCCACGGCGC CCGACCAGTT TTTGACATTT CTAAGCCAAA 5040

AGTTCCATTT GATGAGGTCT TAGATGCAGG GGCAATGTGT CCCTTTTCAG ATTTCAGATG 5100

TTTAGAAAAA GATGTGTCAT ATTTGGGCCA ACTGAAAAAC TCTTGATATG TAGGTTTTTA 5160

TGAAGCTGTG CAGAATGTAG GAAATACATT TTAGAACCAA CAAAGAGGCA TTTAATTTTG 5220

AGTGTGCCTG TCTCCTTTGA GATGAGCAAC AGCTATTTTT CTCTTCAAAA GACAATGCGT 5280

GTATTTATCA GCACATTTTA TATAATCAGC AAATCTAAAC CTCTGAATTA GGTAAGCCCT 5340

ATAGGTTTGT TGCCAGAATT AGTGAATTTA TACATGCAAA GTGCTTAGAA CAGTGCCTGG 5400

TACACAGTGA GCACTCAATA TTATTTATTG CTATTATTAT GTTTATTTAT TTTATACTTT 5460

TAGAGTATAA TTTTGATGTT AGGTTTGGAT TGCTGAGGCC AAGCAAAATT TAGATAGACC 5520

AACCCAGCTA ATCCACTAGA AAGATATTTG AGGGTTATTC CCATCTAAAG ATCTATGGGA 5580

TCTTTGGATA TCATCTGTGA GAAACAACAG AAGTTTGTAG ATAAGACAGA TATAGATTCA 5640

AATGCCACCT TCACAATTTA TTAGTGAGGG AACCCTTTGG TAAAATGAGC ATGACAAAAC 5700

CTACTTCTCA GCTTTGTGAA CGTACAGAAG ATCATGAATG TAAAATGTCC ATGAAGTGGT 5760

AGGTGGTCAA CAGATATTGC TCTAGCAAAG TGGTTAAGAG CAAGCAAACT CTGGAGCCAA 5820

AGGGCCTGGG TTCAAATCCC GATTCTGCCA CTTCTTATAG TATGGCCTTG GGCAGGTGAC 5880

CTCACTTTTC TGTGCCACTA TTCAATGATA ATATTCCTTT ATTGTCCAAC GTTTTGTAAG 5940

TTAAATCAGT TAATAAACAC ACTATGATAA TGTGTTGGTA ACTATTCTTT TTACTTTTAG 6000

CAGAATAACT TAAAGGAACA CTGCAGGTAG GAGGGTTATA CATAATCTCT GAGGGCCAGC 6060

TAGGACACTC GCCCATTTCC CACCCTTTTT CCTGTGCAAT GAAGAGTATA AGAAGTGACA 6120

GTGCCATTCT AAAGGACTAG CCTTGAGTTG GCTCTAATTT ATATGACTCG TGCCTGTAAT 6180

CCCAGCACTT TGGGAGGCCG AGGCTGGCAG ATCACNTTGA GGTCAGGAGT TGGAGACCAG 6240

CCTNGGCAAA CATGACGAAA TCTCATCTCT ACCAAAAATA CAAAGATCAG CCGGGCATGG 6300

TAGCACNCAC CTGTAATCCC AGCTTCTGGG GAGGCTGAGG CAGAAGAATT GTTTGAACCC 6360

AGGGGGCGGA GGATGCAGTG AGCTGAGATC ACGCCACTGC ACTACAGCCT GGGTGACAGA 6420

GTGAGATTCT GTCTCAAATA ATAATAATAA TTTATATGAG AAAGAAGTCA TTCAAAAGCA 6480

TCATTACTTT ACATGTCAAA TTAGAAAGGC ACACCCCAGT ACTAAAGCAT CCTTGATGAT 6540

GAAAACATTT AGAACCACTG GTTTCAGGAG CTCCATGCAA TGGTGAAACA GCCTCTACTC 6600

CAAGGGTTGT TGCTCCCTGT GGGATTCCTG GGTGAGGAAC ACACTGCTCC CGTTGGGGTG 6660

GAATCCTGTG GAGGAAGTGG ATGAAGAGTG TAGCCAAGTC AGTGAGCCTA CTGCATGGGA 6720

TTAGAGTACT TCAGGTTGCA GTATAATTCT GTTCAGGTGC ATGCTCACTC CATCTGGCGT 6780

AAAGAACAGA GAAATTAAAC CATTGATTCA CAGAGCAATA TGAGTAGCTG CCTGGGGACT 6840

TCCTGTCCAT TCTGGGGTNC CCAACAGCCA ATCAATATTG GCCGGTTCCT AATCTGACCT 6900

AGAGCTAGAG GCCACTAGCA CCCCTCCATT CATCCTTTCT TCTCTCCCTA CCCACTCCCA 6960

CCAGCATTCT GAGGAGGTGA GGGCTGAAGC TGCAGAGGCT GTTGTACTGT CAGTTACTGT 7020

AAAGTCCCCA TCCTGCTCTC TGTAGTTTGC TGTGAAGGAG TGGAGGGGGC TGGGAACAAA 7080

GGGGGTTCCC AATAAGCAGC TTACTAATAC CTACCCTTGC TTCTCTCACT TCCTGATCAA 7140

TAATTCCCAT CCTTTCTTCA TGTACCTCCC CCACATTTTT GTTCTTTAAG GGAAGAAGGG 7200

AGCAGCTTTC CATCCCCACA CAATGTTGGG GACATTTGGT ATTATACATT ATGAAAATAA 7260

GATTTACCAG ATTTTAGAGA TGGAAGAAAA CTTGGGGATG ATCTTGTTCC ATTCTCTTAT 7320

AAGAACAAAC AATATTTGGA GAAGCTGAGT AACTTGCTTT TTCAACTTCA CACTTGAAAG 7380

TGTTTTCATG AGGAAGTTGG GGTTCTCTGC AGCACTTGGA TGGGAGTCAG GGACTTGGAT 7440

TGTCCCAATT CTGTCACTAA ATTTGGACAA GCCACTTAAC TTTTCCAGAA TCTAGTTGCC 7500

TCATTTCAAA TATTAGGGAA ATTTCTAAAT GGCTTAAAAG GAGCTTGTTA GCTTTAAAAT 7560

TTTATGATTC TAAGTGTATG CTGCCAGAGA TATGTAGCAT AGCAGGACAC ATTAACAAGA 7620

TTATTGAAAC TGTTCTAATA AAGGACATCT TTGTGTCTTG GGTAGCTACT ATGTTTAAAG 7680

ACTGTGCTAG GTGGGAGTTG TGCAGAATAC ACAGGTTTGC TGTAGAGGGA TAGGGCGTGT 7740

ACACAGACAA CTCTACTAAC AAGAACGTTA CTAGAAGCTT ATTGGAATCA CAGTATTTCT 7800

TGCTGAGGGT ATGAAACATA AGAGTTCTCC TTGGAATATG AGGTTCTATT TGGGGCTTAA 7860

AGAATGGTCA AAGGTTGAGT GCAAATAACA TGGATTGAGA TGGCTTTAAA AAATAATCAA 7920

ATGGTTTGTT AGTATTAAAC TGGTGCAGAA ATAATTGCAG TTTTTGCCAT TCCTTTTAAT 7980

GGCAAAAATT GCAGTTACTT TTAAACCAAA TCCCTAATAT TATTTGCATA GTTTATCTCT 8040

GTTATGGAAG TTTTTATTGA CAAGTAATGT AGATATTCAC CTGATCTAAG TTACCCTGAA 8100

TCTTATATTA GCAGAATCTG AATTGCTTAT AAATAATTAT GGCTATGTTG GATGTAGAAC 8160

TTATTATTTG ATAGTTTATG AACAGTGCTA AGGTCTAATC TACTTTTTAC AGAGAAGCTA 8220

AGAACATGCT ACAGCTGGTT GAAAAACAAA AACTTCAGGC ATTGAAATGT TTTGTCAATG 8280

AAATGGCAGG ACTCATTTGA TGACTGATTA TTATCAACTG ATTTAAATGA CTGAATTTTT 8340

GGTACTGTGT ACATCTATAC TCTAAGAAGG AAATTGAAAG TAATTCTGCT ATGCTTGTTG 8400

CCACTATATT AATAACTGCA TCATCTAAAA TAATTGATAG AGCTCAGATT TATCCTTTGT 8460

AATAATTCTA GTACTTCTTT AAACATGTTT TGGGATTAGC AGCTGTCAAC AGTTAGAACA 8520

TGAAACAGAT TCTGTTACAG GAGTAGAAGT CGATCCAGAC ATTTAATGTC ATTTTCACCT 8580

GTGAGAGAGA GAATAAAGAG AAAGAGAGAT CATTATTTAT GGGATTATGT GAACTTCAAG 8640

TCCGTTTTCA TTATTAGGAG AAGCTGTGCT TTAAAGGACA GTCAGGGACT TTACTTTCAT 8700

GAAATGCCTG AGCTGTAAAT AAAGTATTGC TTTATTTTTT ATTTCTTGAA CATTTGAAAT 8760

AAAAAATTAG CTATGAGTTA TGTTCAAATT ATATTATAAA AATTTGCTCT TAGCATTGTG 8820

CATATATATT ATACAGAAAA ACACAGAGTA AAAAGAATAG ACTTCAGTTC CTGTTCAGAA 8880

AAGGTTTAAA ATTTGAATAC TGATTTTGGA AACCCCAAAC CTTAAGAATT CAAGAAGCTT 8940

ACGGTCTTCT TGAGGGACAC CTATTCAAAC TCTTAAATAT GGTGATTGGG TAGAAAGTGC 9000

AGAAAAGCCT GCTGATACAT GCCCTAAAAC ACCTTGGAAA AAAGAGGTGG TAGTTGCTTG 9060

AGGTAGGACT TAAGTACTAG TTGGAAATAG AAGACAAGGA TGGAGACTGT TGGTAGATGA 9120

CTCTCCATGG GTCCTTCCTG TTTCTACACA CCTTGTAAGC AGGGCATTGA GTGCCTGTGT 9180

TCCAAACTAC CTTTTCCATC ATGTTTCTAC AGCAAACAGT CATGGAAGAT AGAAATAGAG 9240

TCTTCCTCTG GAGCAAAGGG CAGACACGCT TGCTTCCTGT ACTTCCCACT ATAAGATATT 9300

CCGGCTCCCT AAACTCAGCT GCCTTTCCTG TAACCCACCA TGATACAGAT GTCACCTGAC 9360

CTGTGGGAAT TGGGGGTCAG GGAACCAAGA GAAATGCTGA CTGTCTGGCT ACTGTGACTG 9420

CCCTGAGTAA TAAATTGTCC TTCGTCTCCA ACCCAGGAGT CTCATGTTTT CTACCAGCAG 9480

GATAACTGTG GCGGGCTAAC GTGTTAGTTT GCAAGTAAGG TAAAATCTCA GACCCTTTGC 9540

AGTTTGTGGC AGGGATTATA TTCTGAGGAG AGAGGAACCG TATGCACCAT GGCTCAGAGG 9600

CATGAGAAAC GGGGAACCAT AACTAGTTCT CTATCTTCAG AGCCTTTAAA AGGTGCACCA 9660

AGGAGGGCAT TTTAGGGGAG AATATAAAGT TGGAGATATA GACACAGCCA GATTCCTGAG 9720

AGACCTTATA TGCCAGGTAG AAGACTTCAG ATTGTATGGG GGAATTATTA GAGAATTTTT 9780

AGCAGGGGTG TGATATGATA AATTTTGTGT TGATTAAGTT ACTCCAGGAA ATATGCGATG 9840

GGTGGATTGA AGGATGGGGC ACCTTTTNTC TAGGACGAAA AAGAAAGAGT AGTTGGTGAA 9900

GTCAGTTAGA GGTAGTAATA GGATGAAGAA GGGATCTGAA TGACCCCTTG GCCATTCAGT 9960

GAGTAGTGAT GCTATTCACC TAGATACAGC ACATAGNNGG GAAANGAAAT NCTGGGAAGG 10020

AGGGAGATGA GACCGAGTTA GCTTTAAAAT AACTAAATTC AGGCCTAGGA GCCTATAGGC 10080

TATCCAGATA GAAATATTTA ATNGCCTATA TGGATCTGGA ACTCAGGAAG GAGGCTTCCG 10140

TGGGAGCAGA ACACTTGGGC ACCATTAGGG TGTATGTGGT ANATGCATTC TTGTGCAGCA 10200

GTCAAGGGGA TGGGATTTAG ACTCAAGTGC AAATTGCCCC CCATCTCCTG TGATAAGTGA 10260

CTGAAGCTCT CCGGGCTTCA GTTTCCTAGT TCATCATAGT GGGCTCTAGC GGATAAATGT 10320

TACAAAGGTT AAATGAGACA ACATAGGCAA AGTGCGTGGT ACTCAATAGA AGTCAGCTGC 10380

TGTCATCAGC AGCAGGATCA CCAGAATGTG GTGCTTGACA CCAAAAGATT AGGTGAGATT 10440

GCCCAAAACA GCAGGTGAAA TGAGGGGAGA GGATGNAAGT CAAACACAGG AAGAAAAGCC 10500

TTTGAAGTAT GTGGAAAGAA ACAACCAGAA AGGTAAGATA AGAACCAGAA GAGATTCAAG 10560

AAGGAAGGTG TGGCCGGGCG CGGTGGCTCA AGCCTGTAAT CCCAGCACTT TGGGAGGCCG 10620

AGGCGGGCGG AACACGAGGT CAGGAGATCG AGACCATCCT GGCTAACACG GTGAAACCCC 10680

GTCTGTACTA AAGATACAAA AGAATTAGCC GGGCGCGGTG GCAGGCGCCT GTAGTCCCAG 10740

CTACTCGGGA GGCTGACGCG GGAGAATGGC GCGAACCCGG GAGGCGGAGC TTGCAGTGAG 10800

CCGAGATCGC GCCACTGCAC NTCNAGCCTG GGCGACAGAG CGAGGAGCCG TCTCAAAAAA 10860

AAAAGAAAAA AAAAAAAAAG TAAGGAAGGT GTGGCCAAGA TTGAGAAATT CGTCAGAGCA 10920

AACAAGGCAG TCAGGGGCTA AATAGCCTCC TTTAAATTTT ACAACCTTGA GGACCTCGGC 10980

AACTTTAACA GAATTTCAGT GGATCCCTAG GGCAAACCAG GCCTTACAAA CCAGGAATGG 11040

ATGGTCAATA GGAAGTGGAG ACAGTAAGTG TAGACCTTAC CTTGGAGGGA AGGNAAGAGA 11100

AAGAGCCATG GCCAAGGGAA GTTTGAAATC AAAGGAAATA TCTTTTTTTT TTTTTTCGA 11160

TTGGAGAGAC CTCAGTTATT CTTTTAAAAT ACTTATTGAG CCCCTCAGTT ATTCTTTTAA 11220

AATACGTATT GAGTCCCTAC TTTGAGTCAG GCACNATGGC AGACACGAGG GNGATAGCAG 11280

TGAATCAGAC AGATGCAACG CCTGCCTTCA TGGAGTTTCA CCTTAGCATC TGTCCATATG 11340

CTAGGGGAGT GGGGCAGGGG CAGGGAGCTG GATACAGGAG AGACTGAAGA TCCAGGGAGC 11400

AAGTGAGTAA AGAATAGGGC TTGAGATCCC ACAGACAACT CAGCTTTGAA CAAAAGGGTT 11460

TTGTCATCCA ATAGGACAAG AAGGCGTTAG GATACATCAA ACGTGGTTGT TGAAAACAGA 11520

AAAGGGCTGG GCACTGTGGC TCATGCCTAT AATCCCAGCA CTTTGGGAGG CCAAGGTGGG 11580

CAGATCACTT GAGGCCAGGA GTTCGAGACC AGCCTGGCCA ACATGGTGAA ACCCCATCTC 11640

TACTAAAAAT ACAAAAATTA GCCAGGTGTG GTGGTGCATG CCTGTAATCC CAGCTACTTG 11700

GAAGGCTGAG GCAGGAGAAT TGCTTGAACC CAGGGGGTGG AGGTTGCAGT GAGCCACGAT 11760

CGTGCCACTG CACTCCAGCC CGGGCAACAG AGCGAGACTC TGTCTCAAAA AAAAAAAAAG 11820

GAAGAAAGAA CATAGACAGG GAAATGTAGT TAAGGNNAGT TTGGGTTTGG GTTTGGTAGA 11880

AGCGTTTTCT GTTNNTTGTT TGTTTGTTTT CAGAAAGAGT CTCACTCTGT TGTCCAGACT 11940

GGAGTGCAGT GGCACAATCT TGGCTTGCTG CAGCCTCTGC CTCCTGGATT CAAGCAATTC 12000

TCCTGCCTCA GCCTCCTGAG TAGCTGGGAT TACAGACACC TACCACCACA CCAGGCTAAT 12060

TTTTGTATTT TTAGTAGAGA CGGGGTTTCA CCATGTTGGC CAGGCTGGTC TCAAACTCCT 12120

GACCTCAGGT GATCCACCTA TCTTGGCCTC TCAAAGTGCT GGGATTACAG GTGTGAGCCA 12180

CTGCACCTGG CCTAACATTG ATATCTGTTG ATGAGAAGAA GCCAGGTGTT GGAGTGATAG 12240

CTTATAGCAC ATGAACTGAA TAAAACAGTG TTTAAGACAA TGTTTGCAAC ATAATAGGCA 12300

CTGAAGACAT GTTAATGGAA GGTGGATTTG TGATTCAGAA CCTCTAGACT ACCTGGGCGA 12360

GTCTTTTAAA ATGTAAGTAA TATCTTAAGT GATATTACTT GTCCCAGATC AGTTGTTTAA 12420

AACTGAGGTT TAATGCTGTC AGAGTAGCAC TGTATCGTCT TCTATCATGG GGGCCTTTGT 12480

TGGCTTTAGG AGGTTTGTGT TTCATAGTAG TTTCCCAGTG GGCTCTTTGT TACCTGTAAT 12540

GAGTGTGACA GTTATGCCAT AACCAGGTTT TATATGGAAT ACAATTTTGA GAAAGTTCTT 12600

TCTAGGCAGA GAAGCTTATT TGAACCTCTT ATTATATTTG GGTTTCAGGC TTTTGAGTTC 12660

TTCTGAAATA ATAGCCCTTT GAAGGTAGCT ATTGCTATGA CTTCATTAAA TTCTAATGCC 12720

TCTGGTTTTC TCCCCCAGGT TTCTGCATAT GAAGTGTGTA AAATAGATTG CTTGATCCAA 12780

AACAGAAAAA CAGTGATAAC TGTTTTGCTG AGTTCCCAGA CCCTTCCCAA GATGGAACCA 12840

ATAACATTCA CAGCAAGGAA ACATCTGCTT TCTAACGAGG TCTCGGTGGA TTTTGGCCTG 12900

CAGCTGGTGG GCTCCCTGCC TGTGCATTCC CTGACCACCA TGCCCATGCT GCCCTGGGTT 12960

GTGGCTGAGG TGCGAAGACT CAGCAGGCAG TCCACCAGAA AGGAACCTGT AACCAAGCAA 13020

GTCCGGCTTT GCGTTTCACC CTCTGGACTG AGATGTGAAC CTGAGCCAGG GAGAAGTCAA 13080

CAGTGGGATC CCCTGATCTA TTCCAGCATC TTTGAGTGCA AGCCTCAGCG TGTTCACAAA 13140

CTGATTCACA ACAGTCATGA CCCAAGTTAC TTTGCTTGTC TGATTAAGGA AGACGCTGTC 13200

CACCGGCAGA GTATCTGCTA TGTGTTCAAA GCCGATGATC AAACAAAAGT AAGTGAGATG 13260

GAGATCCAAA AGACTAAGGT GTGGCTGGCT GGTTTTTATT GTATGGGGGT CAGGATATTT 13320

ATTTTAAGTA TACTGAAATG AATAAGGAAT TAATGCTGCA GTTATAAATT GATTACTTAG 13380

CTGAATTTTT GTTTTATGGT GATAGTTTAT AGTTTTAAAG CACATTTGAA AACAGATACG 13440

AGAAATTATC AGTTTTTGAG TTCAAAAATT CAAGAGAAAT CAGTCTAAAA CTACTAATTA 13500

AGAGCAGAAG TGTTAAGATG TACATTATTT CAGATGAATG TTCTAAAGCC ATGCCTCTCA 13560

AACTGAAATG AGCTTGTGAG TCACCTGGGG ATCTTGTTAA AATGTGAATC TTGATTCAGT 13620

AGGTCTGGGG TGGACCCCAA GACTGCATTT GTAACAAGCT GCCAAGAAAT GCTGATGCTG 13680

CCCTTTTGCA GGTTGCACTT TGAGTGGCAA AGTTCTAAAT CTCCACATTT GTAATCCTAT 13740

TAAGAAAAAT ATAGTCATTC GTAAACTGTG TAAAAATGCT ACTGGCCAGT TTCCCAAGGC 13800

ATAATGTTCA CTTAGGCAAA GGTCATTGAT AAGAACGCTG GATATGCATC TAAGTTTTGA 13860

TGCGATCAGG GGTTCTTTGT GTTTTTTTCT TTCGCAAACC TCAGGTCAGA TCTGATTAGC 13920

TTGTTATTAT CACATGATAT GGCTGAAAAA AAATGTGAGA CATGGTAAAA GTTCTGCTCT 13980

TTCCTCGTTC ATTTGTGCTT GCTTTGTTAT TAGCATTCGT TGTAGCTCTG GGCAGGACTC 14040

ATTTGAAGAT GCTTGNCCCA TTTTATGAGG ATTAGCTTAG ATAAAATTGA AAATATAATG 14100

CAAATAGCAA CTTTCTCAGT TGGGCTCAGG GCTCCACAGC TAACCCCATG GACTGTGGAG 14160

TCTTGCCGTT GTTTTGGGTG CCAAGCAAGC CAAGTCACAT GTGATTCAAG CTGTCTGCCA 14220

CATGTACAGG GCGAGGATGC GAGTGTCAAT CCACCTGTTA ACTGTCAGTG AAGCCTTGAA 14280

AGCTTCTCAT ATTTTCAAGG TTAAAATCTG GATAGAAATG CTAAAGTTTT CTCTCTGCAC 14340

TCCATTAGGT TATTTTATGT ACTCTCTAGG GTGTAAGGAC CTTATTTAGA AATTAATATT 14400

CTTGGTATCA AGTAGATGCC CTTTTGCTTG TTCATTTGTT GGTTCTTCTA GTCATTCAGA 14460

ATTGCTGTTG CAGGTACTGT TGGAGATGAT ATTAGCAGAG GCTTGTAGGA AGGCAGGAGC 14520

ATCAGTGGGG AATAGGACCA GGTGATCTAT GTATAGGACA TAATGGAAGG ACTGAGAAGG 14580

GAGCCTAACA CACACCCAAA GGGTAGAGAA GGCTTTGTGA AATAAAGGCT AATATGGAGC 14640

TCAAAACCAC CATTTCACTC ACAGAATCAA ACTCTCATAT TATAAATCAT TTCATGTTAT 14700

TGTCCACACA TCTCAAGTGG GCACGGCAGC ATCAGGCTTG GAGATTCAGA GGGACTAACT 14760

TCCTGTACTC TAATCCTACT TCTGCACCCA TAAACTGGGT GGCCTCAGGC AATTGAGTCT 14820

GTTTTCTTAT CTGTTAAATG GGGATAATTA CAGTATTTAT CCAATAGAGT TGCTGGAAAG 14880

ACTAAATGAG GTAGCACTCG ACCTGAAACT TAGTAAGCAT TTATAGCCAT AAAAACATTT 14940

TCATTCAAGA AAATTTTACT AGAGGCAGAT TATATGCTAA TTTCATTTCA CGTCTTAGGT 15000

AAAAAGAAAC ATGATACCTA GATGAGTGCC TTCAGCTTTC AAAGATGAGA TTCTGGTCAT 15060

ATTTGAGGAA CATTTTAAAA ACTACACGTA TAACTTAATG GCTCCTATTA TTTGGACAAA 15120

TTCCAGAATG AAAATGAGAG GACTGAACAG CCTGTACCTC AGTCCAGCTC TATATAGTAT 15180

TTGGACTGAA TTTCCTTGGG GAGAGTTTGT GCGTGGAATC GTTGTTCAGC ATTTTACACA 15240

TTTGACTCTT TCCCAAAATC TTTTACGGCC ATCTGAGAAT AGGCTTCTGG CCAGTCATTC 15300

GGATGCCTGA CAAGAGAAAG AGATTTATAA CCAAATTCTG TAATTGGGAC TTCCAGTCTT 15360

TCCCCAAGTA GAGAATTGGA CTTACTCTAT ATGCTAAAAA CCCATGGTTG AAATATGAAT 15420

TAGTTCTTAA GTGATTTTTG GCTTGCATAC CATTTTTGCA AACACAAATT GTCATTACTC 15480

TGCTCATTTA ATAAAAGAAT AATTTGTAGT ATAGGTATAT ACCTCAATCA GTGATTTTGT 15540

TGTTGGAAAC AGAACAGTAA ATCACACTGG CCATGATGCT AACAGCGTGA TAGATTTTCT 15600

GTTCTTGGGA CACCAATGTC ACTGTATCTC ATAGCGAAGG ATTATCTGCT GTAGGAGCAT 15660

TCTCTTGACT ACTTATAACA TTTGCTGGGT GAAATAATTC TCCAGGTTAA GGCCTCTTCT 15720

AAACAGATGA GGTCAGCACT AACTGCATTT GCCAGAGAAG ACATATGCAT TTACTGCCAG 15780

CATCATAAAC ACAAAACTAC AGTTTGCGAG GAAACCCTTT GACCAGCATC TAATTAATTC 15840

ACTGAGTAAT GTCTTGGGAG AAGAGGCATG TAAAGGAACA ATTTTATAAG CATGCCATGA 15900

GATTGTTTTC NNATTGTATG TTCCATAGAA TATGAGGAAA CTTCAAAACA TTTTGTGGAA 15960

AAATTGAATT AAAAAGNNAA AAACACATAT ATACATAAGC TTTATTTCTC AAGATAAACT 16020

TTATCAAGTT CAAGACACTT TTGTAAGCAA TGTTAACAGC CATTGAGTCG GTCTCTAAAG 16080

AACTGAGGGT CCTGGGAATT TAACCATGTT TATACAGTCT TTTATACATT ATTAACTGGA 16140

GAAAAATTGG CGCTCTTTAA AGATTTTTTA AAATTGAGAA GCAAAAGGAC GTCAGAAGGA 16200

GCCAAATTAG GCCTGNNAAG TGGATGCCTA ATGATTTCCC ATGGAAACTC TTGCAAAATT 16260

GCTCCTGTTT GATGAGAGGA ATGAGCAGGA ACATTGTCAT GGTGGACAAG GACTCTGGTG 16320

AAGCTTTNNC AGGCGATTTT CTGCTAAAGC TTTGGCTAAC TTTCTCAAAA CACTCTCATG 16380

ATAAACAGAT GTTATCATTC TTTGGCCCTC CAGAAAGTCA ACAAACAAAA TGCCTTGGGC 16440

ATCCCAAAAA ACTATTGCAA CCATTTGCCC TTGACCAGTC CACTTTCGCT TTGACTGGAC 16500

CACTTCTGCT CTCAGTAGCC ATTGCTTAAA TTTGTCTTGA TCTTTAGGAT TGCGCTGGTA 16560

AAACTATGTT TCATCACCTG TTACAATTCT TTGAAGAAAT GCTTCAGGAT CTTGATCCCA 16620

CCCGTTTAAA ATTTCCATTA GAAACTCTGC TCTTGTCTGC AGCTGATCTG AGGGCAATGG 16680

TTTTGGCACC CATCTAGTAA AACGTTTGCT CAGTGTTAAT TTTTCATCCA GGATTGTGTA 16740

AGCTGAACCA GCAGAGATGT CTATGATATT GGCTAGTTGG TCCTCTTCAA TGAGGGCATG 16800

AACAAGATGA ATATTTTCCT CAAACAATTA TCTGGATGGT CTGCTGCTGC AGGCTTCATC 16860

TTCAATATTG TCTCGTCCCT TCTTTTTCTT TTCCCCCCCG CTTGAGACAC AGTCTTGTTC 16920

TGTTGCCCAG GTTGGAGTGC AGTGGCCCGA TTTCGGCTCA CTGCAACCTC TGCCTCCCGG 16980

GCTCAAGCGA TTCTCCTGCC TCAGCCCACC AAGTAGCTGG GATTACAGGT ACACATGATC 17040

GTGCCTGGCT AATTTTTGTA TTTTTAGTAG AGACAGGGTT TCACCGTGTT GGCCAGGCTG 17100

GTCTCGAACT CCTGACCTCA AGTAATCCAC CTGCCTTGGC CTCCCAAAGT GCTGGGATTA 17160

TAAACATGAG CCACCACACC TGGCCTCATC CTTTCTTAAA ATGAGTTATA CATTTGTAAG 17220

CTGCTGATTT CTTTGGACAT TGTGCCTATA AACTTTTTGT AAAGCATCAG TGATTTCACC 17280

ATTCTTCCAC CCAAACTTCA CCATAAGTTT GATGTTTCTT CTTGCTTTGA TTTTAGCAGG 17340

ATTCATGTTT CTCTGATAGG GGGTCTTTTC AAACTGATGT CTTATCCTTC TTAGAGCCTC 17400

ATCCCAGATC CTGTTCAGAC ATGCTACAAG TTAATACAAG TTTATTTGGT GCCAAAAAAT 17460

GGAAATCCAT GCATAGTTTT TAAATAATAT GCATTTTTCA TGNACTTTTT GAAGACCCCT 17520

TGTATACTTA AACTGCTCCA CATGGAAAAG CTTCCATGAT CAAATGCAGT AAGGCAGCAT 17580

CTCAAACAT 17589

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1999 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GGACAGTATC TAGCACAATA CCCCAAATCG ACTAACTCCT CCGTAAAGAA TAGCTACCAC 60

TATTGTGAGA GTTTTAAGTC AAGCTGTGAA TAAAACTCTT GGGTCCACTT AAAAATACCT 120

CCCCTGGATG TAAGCATCCA GGGAAATCAG GGAATGCCAT AAGACAGCCC TAATCTAAAA 180

GCCTACAAGA AGCTCAGTGG GCTTCAAGGA AGACACTGCT CTTGGTACGA TGAGGAAACC 240

TGGCCCTCTA TTTGCCTCCT GGGCCACAGT AATATTGATA ATAGCTGCTG CTTTTAGTTG 300

AGGACCATGT ACGTCTGTGT CACTGCACTG GCCACTTTAC TTACACTTTC CTGCTTTGTC 360

CTCACAAAGA TCCTGTAAGG TGTGTATTGG TCCCATTTAG CAGGTAAGAC AATGAAGACC 420

AGAGGTCCAG CACCTTGCCT AAACCACACC TGCTGGGATT TGGATTCAAG TCCAACCGTA 480

CAGCTCAAAC GCTCAGCCAC TTCCCTAAAG TCCACCCCCA GCTACATTAA GTAAAAAAAT 540

CCAGAAAGAT GCCACCTGGG GGTCTGGAAC TGCCTCCTCC GAGCACCCGG CTCTCCCCTC 600

CCTGCGGACT CTTCTCTGGA GAGGATGTGA TGCTTCTTAC TTTTCTCAGA TCCCTCTCCC 660

CACCCTGCGA GTGACGTTGC GCCTCTGTGC CTGGTGGGAT AGGGATCTGG GAGCTTCGCC 720

TGTTTTTTGC ACACTGCCAT CCCCTAGTCT TAGGGAGCGA GCTCTGTCCC GCTTTTCACA 780

TCTCCGCGTC TTTCCTTGCA CTCTACATCA CCGCTGGGAA TGTCCCCAGA CCTGATCGGG 840

GCATGCACAC TGGGGTGTGC GTGTGCGTGT GGTGTGTGTT CCTGCGCGTG TGCCGGGCTC 900

GCGGGGCAGG AAAAAGCGCC TAATCCAGGC TCTGCGTCAC TCCCGCAATT GGTTAGAAAT 960

GGAGTTTCCT GGTGTTTAAT CCCGGGAGGG CACTTCGCCT TCGTTGTTTC CCAGAGTCCC 1020

TGATTTTCCT GCCTCGCATG CCAGCGCCCC ATAGGGCATC CGTGCCTCAG TTCACCTCTT 1080

GCCATCCTCC AAGGACGGGG AGAAGGGGTA AGGCGGGGGA GAGCAAGGTG GCTTGGTCGC 1140

CCCCGGCCCC CGCCCCCCAT GTTGTGTGCA GTTTCCACCA CGTCTGTTTC GGAGGGAGAA 1200

GAGGAGGGTG CAGATGAGGC GAGGCGCCTT CGGGAGCGCG GAGAGCGGGC AGGCAGTGCC 1260

ACCTGCTGAG AGCCACTCAG GCCGAGCAAG CGGCGGGCAG TGCCACCTGC TATAAATAGG 1320

CCGCCAAGGA CAGGGTGTGC GACTGTACAT CCCGCCACGA GGGCCTGCAT CACGCGCGGG 1380

GCCCCGCGCC CCCGGCTCCC CAGGGAAACG CTGTGCCCAG ATCCTGCGCA GGGGTCTGGA 1440

TGGGGCGGCG GCCCGAGTAC TTCCCCCCTA TTCCCCCCAC ANACACTGGC TGAGGATGGC 1500

CCGCGGGCTT GGGGGCGGGG GGTGGCAAGG AGGGGAGGGA GGCCGCGGCG GACCCGCAGT 1560

GCAGCAGCTG TTGCTCGCGT GTGACTCGCC CGTCCGGGCC GTGCTGCCCA GGCACAGTCA 1620

CACGGCGCAG TGGGGAGGAG GAGGACACCG AGTCCCCCTC CCAGCTCCCC GGGGACCGAG 1680

TGGGGAGATC CCGGCTCCTG TCTTCCCCTC GCCTCCAGCG CGCTCGCCCA GGCTGGGAGG 1740

AGGAAACCAG AGCCGCGCGC AGACACCTCC TCCTTCTCCT CCTCTTCTTC CTCCTCCTCC 1800

TCCTCCTCCT CCTCCTCTTC GGCTGCTGCT CCTGGTGCCG CCACCGTCCG CCGGTGCCTG 1860

TTGCTGCCGC CGCCGCGGGA CCTGCTGTGT CCTCAGCTGG GTGGAGAAGA GGCGGGCGCC 1920

GAGCCGAGGG GAGCCCCCTC CCCGTCCCCC CGCGGCGGGA AGAGCGCAGC CAGCCGGGTG 1980

CGATGGACTC CCCGCCCGC 1999

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3984 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CCAGGCCGTC CCCAGGATGC CCCCAAGCAC CTGCGNGTCC CGGCCCGGCC CCGGGCTCTG 60

AGCGCGCCGC GGCACAGGTT TCTGCATATG AAGTGTGTAA AATAGATTGC TTGATCCAAA 120

ACAGAAAAAC AGTGATAACT GTTTTGCTGA GTTCCCAGAC CCTTCCCAAG ATG GAA 176
Met Glu
1

CCA ATA ACA TTC ACA GCA AGG AAA CAT CTG CTT CCT AAC GAG GTC TCG 224
Pro Ile Thr Phe Thr Ala Arg Lys His Leu Leu Pro Asn Glu Val Ser
5 10 15

GTG GAT TTT GGC CTG CAG CTG GTG GGC TCC CTG CCT GTG CAT TCC CTG 272
Val Asp Phe Gly Leu Gln Leu Val Gly Ser Leu Pro Val His Ser Leu
20 25 30

ACC ACC ATG CCC ATG CTG CCC TGG GTT GTG GCT GAG GTG CGA AGA CTC 320
Thr Thr Met Pro Met Leu Pro Trp Val Val Ala Glu Val Arg Arg Leu
35 40 45 50

```
AGC AGG CAG TCC ACC AGA AAG GAA CCT GTA ACC AAG CAA GTC CGG CTT      368
Ser Arg Gln Ser Thr Arg Lys Glu Pro Val Thr Lys Gln Val Arg Leu
                55                  60                  65

TGC GTT TCA CCC TCT GGA CTG AGA TGT GAA CCT GAG CCA GGG AGA AGT      416
Cys Val Ser Pro Ser Gly Leu Arg Cys Glu Pro Glu Pro Gly Arg Ser
            70                  75                  80

CAA CAG TGG GAT CCC CTG ATC TAT TCC AGC ATC TTT GAG TGC AAG CCT      464
Gln Gln Trp Asp Pro Leu Ile Tyr Ser Ser Ile Phe Glu Cys Lys Pro
        85                  90                  95

CAG CGT GTT CAC AAA CTG ATT CAC AAC AGT CAT GAC CCA AGT TAC TTT      512
Gln Arg Val His Lys Leu Ile His Asn Ser His Asp Pro Ser Tyr Phe
    100                 105                 110

GCT TGT CTG ATT AAG GAA GAC GCT GTC CAC CGG CAG AGT ATC TGC TAT      560
Ala Cys Leu Ile Lys Glu Asp Ala Val His Arg Gln Ser Ile Cys Tyr
115                 120                 125                 130

GTG TTC AAA GCC GAT GAT CAA ACA AAA GTG CCT GAG ATC ATC AGC TCC      608
Val Phe Lys Ala Asp Asp Gln Thr Lys Val Pro Glu Ile Ile Ser Ser
                135                 140                 145

ATC CGT CAG GCG GGG AAG ATC GCC CGG CAG GAG GAG CTG CAC TGC CCG      656
Ile Arg Gln Ala Gly Lys Ile Ala Arg Gln Glu Glu Leu His Cys Pro
            150                 155                 160

TCC GAG TTC GAC GAC ACG TTT TCC AAG AAG TTC GAG GTG CTC TTC TGC      704
Ser Glu Phe Asp Asp Thr Phe Ser Lys Lys Phe Glu Val Leu Phe Cys
        165                 170                 175

GGC CGC GTG ACG GTG GCG CAC AAG AAG GCT CCG CCG GCC CTG ATC GAC      752
Gly Arg Val Thr Val Ala His Lys Lys Ala Pro Pro Ala Leu Ile Asp
    180                 185                 190

GAG TGC ATC GAG AAG TTC AAT CAC GTC AGC GGC AGC CGG GGG TCC GAG      800
Glu Cys Ile Glu Lys Phe Asn His Val Ser Gly Ser Arg Gly Ser Glu
195                 200                 205                 210

AGC CCC CGC CCC AAC CCG CCC CAT GCC GCG CCC ACA GGG AGC CAG GAG      848
Ser Pro Arg Pro Asn Pro Pro His Ala Ala Pro Thr Gly Ser Gln Glu
                215                 220                 225

CCT GTG CGC AGG CCC ATG CGC AAG TCC TTC TCC CAG CCC GGC CTG CGC      896
Pro Val Arg Arg Pro Met Arg Lys Ser Phe Ser Gln Pro Gly Leu Arg
            230                 235                 240

TCG CTG GCC TTT AGG AAG GAG CTG CAG GAT GGG GGC CTC CGA AGC AGC      944
Ser Leu Ala Phe Arg Lys Glu Leu Gln Asp Gly Gly Leu Arg Ser Ser
        245                 250                 255
```

```
GGC TTC TTC AGC TCC TTC GAG GAG AGC GAC ATT GAG AAC CAC CTC ATT      992
Gly Phe Phe Ser Ser Phe Glu Glu Ser Asp Ile Glu Asn His Leu Ile
    260                 265                 270

AGC GGA CAC AAT ATT GTG CAG CCC ACA GAT ATC GAG GAA AAT CGA ACT     1040
Ser Gly His Asn Ile Val Gln Pro Thr Asp Ile Glu Glu Asn Arg Thr
275                 280                 285                 290

ATG CTC TTC ACG ATT GGC CAG TCT GAA GTT TAC CTC ATC AGT CCT GAC     1088
Met Leu Phe Thr Ile Gly Gln Ser Glu Val Tyr Leu Ile Ser Pro Asp
                295                 300                 305

ACC AAA AAA ATA GCA TTG GAG AAA AAT TTT AAG GAG ATA TCC TTT TGC     1136
Thr Lys Lys Ile Ala Leu Glu Lys Asn Phe Lys Glu Ile Ser Phe Cys
            310                 315                 320

TCT CAG GGC ATC AGA CAC GTG GAC CAC TTT GGG TTT ATC TGT CGG GAG     1184
Ser Gln Gly Ile Arg His Val Asp His Phe Gly Phe Ile Cys Arg Glu
        325                 330                 335

TCT TCC GGA GGT GGC GGC TTT CAT TTT GTC TGT TAC GTG TTT CAG TGC     1232
Ser Ser Gly Gly Gly Gly Phe His Phe Val Cys Tyr Val Phe Gln Cys
    340                 345                 350

ACA AAT GAG GCT CTG GTT GAT GAA ATT ATG ATG ACC CTG AAA CAG GCC     1280
Thr Asn Glu Ala Leu Val Asp Glu Ile Met Met Thr Leu Lys Gln Ala
355                 360                 365                 370

TTC ACG GTG GCC GCA GTG CAG CAG ACA GCT AAG GCG CCA GCC CAG CTG     1328
Phe Thr Val Ala Ala Val Gln Gln Thr Ala Lys Ala Pro Ala Gln Leu
                375                 380                 385

TGT GAG GGC TGC CCC CTG CAA AGC CTG CAC AAG CTC TGT GAG AGG ATA     1376
Cys Glu Gly Cys Pro Leu Gln Ser Leu His Lys Leu Cys Glu Arg Ile
            390                 395                 400

GAG GGA ATG AAT TCT TCC AAA ACA AAA CTA GAA CTG CAA AAG CAC CTG     1424
Glu Gly Met Asn Ser Ser Lys Thr Lys Leu Glu Leu Gln Lys His Leu
        405                 410                 415

ACG ACA TTA ACC AAT CAG GAG CAG GCG ACT ATT TTT GAA GAG GTT CAG     1472
Thr Thr Leu Thr Asn Gln Glu Gln Ala Thr Ile Phe Glu Glu Val Gln
    420                 425                 430

AAA TTG AGA CCG AGA AAT GAG CAG CGA GAG AAT GAA TTG ATT ATT TCT     1520
Lys Leu Arg Pro Arg Asn Glu Gln Arg Glu Asn Glu Leu Ile Ile Ser
435                 440                 445                 450

TTT CTG AGA TGT TTA TAT GAA GAG AAA CAG AAA GAA CAC ATC CAT ATT     1568
Phe Leu Arg Cys Leu Tyr Glu Glu Lys Gln Lys Glu His Ile His Ile
                455                 460                 465

GGG GAG ATG AAG CAG ACA TCG CAG ATG GCA GCA GAG AAT ATT GGA AGT     1616
```

```
Gly Glu Met Lys Gln Thr Ser Gln Met Ala Ala Glu Asn Ile Gly Ser
            470                 475                 480

GAA TTA CCA CCC AGT GCC ACT CGA TTT AGG CTA GAT ATG CTG AAA AAC      1664
Glu Leu Pro Pro Ser Ala Thr Arg Phe Arg Leu Asp Met Leu Lys Asn
        485                 490                 495

AAA GCA AAG AGA TCT TTA ACA GAG TCT TTA GAA AGT ATT TTG TCC CGG      1712
Lys Ala Lys Arg Ser Leu Thr Glu Ser Leu Glu Ser Ile Leu Ser Arg
    500                 505                 510

GGT AAT AAA GCC AGA GGC CTG CAG GAA CAC TCC ATC AGT GTG GAT CTG      1760
Gly Asn Lys Ala Arg Gly Leu Gln Glu His Ser Ile Ser Val Asp Leu
515                 520                 525                 530

GAT AGC TCC CTG TCT AGT ACA TTA AGT AAC ACC AGC AAA GAG CCA TCT      1808
Asp Ser Ser Leu Ser Ser Thr Leu Ser Asn Thr Ser Lys Glu Pro Ser
                535                 540                 545

GTG TGT GAA AAG GAG GCC TTG CCC ATC TCT GAG AGC TCC TTT AAG CTC      1856
Val Cys Glu Lys Glu Ala Leu Pro Ile Ser Glu Ser Ser Phe Lys Leu
            550                 555                 560

CTC GGC TCC TCG GAG GAC CTG TCC AGT GAC TCG GAG AGT CAT CTC CCA      1904
Leu Gly Ser Ser Glu Asp Leu Ser Ser Asp Ser Glu Ser His Leu Pro
        565                 570                 575

GAA GAG CCA GCT CCG CTG TCG CCC CAG CAG GCC TTC AGG AGG CGA GCA      1952
Glu Glu Pro Ala Pro Leu Ser Pro Gln Gln Ala Phe Arg Arg Arg Ala
    580                 585                 590

AAC ACC CTG AGT CAC TTC CCC ATC GAA TGC CAG GAA CCT CCA CAA CCT      2000
Asn Thr Leu Ser His Phe Pro Ile Glu Cys Gln Glu Pro Pro Gln Pro
595                 600                 605                 610

GCC CGG GGG TCC CCG GGG GTT TCG CAA AGG AAA CTT ATG AGG TAT CAC      2048
Ala Arg Gly Ser Pro Gly Val Ser Gln Arg Lys Leu Met Arg Tyr His
                615                 620                 625

TCA GTG AGC ACA GAG ACG CCT CAT GAA CGA AAG GAC TTT GAA TCC AAA      2096
Ser Val Ser Thr Glu Thr Pro His Glu Arg Lys Asp Phe Glu Ser Lys
            630                 635                 640

GCA AAC CAT CTT GGT GAT TCT GGT GGG ACT CCT GTG AAG ACC CGG AGG      2144
Ala Asn His Leu Gly Asp Ser Gly Gly Thr Pro Val Lys Thr Arg Arg
        645                 650                 655

CAT TCC TGG AGG CAG CAG ATA TTC CTC CGA GTA GCC ACC CCG CAG AAG      2192
His Ser Trp Arg Gln Gln Ile Phe Leu Arg Val Ala Thr Pro Gln Lys
    660                 665                 670
```

GCG TGC GAT TCT TCC AGC AGA TAT GAA GAT TAT TCA GAG CTG GGA GAG 2240
Ala Cys Asp Ser Ser Ser Arg Tyr Glu Asp Tyr Ser Glu Leu Gly Glu
675 680 685 690

CTT CCC CCA CGA TCT CCT TTA GAA CCA GTT TGT GAA GAT GGG CCC TTT 2288
Leu Pro Pro Arg Ser Pro Leu Glu Pro Val Cys Glu Asp Gly Pro Phe
695 700 705

GGC CCC CCA CCA GAG GAA AAG AAA AGG ACA TCT CGT GAG CTC CGA GAG 2336
Gly Pro Pro Pro Glu Glu Lys Lys Arg Thr Ser Arg Glu Leu Arg Glu
710 715 720

CTG TGG CAA AAG GCT ATT CTT CAA CAG ATA CTG CTG CTT AGA ATG GAG 2384
Leu Trp Gln Lys Ala Ile Leu Gln Gln Ile Leu Leu Leu Arg Met Glu
725 730 735

AAG GAA AAT CAG AAG CTC CAA GCC TCT GAA AAT GAT TTG CTG AAC AAG 2432
Lys Glu Asn Gln Lys Leu Gln Ala Ser Glu Asn Asp Leu Leu Asn Lys
740 745 750

CGC CTG AAG CTC GAT TAT GAA GAA ATT ACT CCC TGT CTT AAA GAA GTA 2480
Arg Leu Lys Leu Asp Tyr Glu Glu Ile Thr Pro Cys Leu Lys Glu Val
755 760 765 770

ACT ACA GTG TGG GAA AAG ATG CTT AGC ACT CCA GGA AGA TCA AAA ATT 2528
Thr Thr Val Trp Glu Lys Met Leu Ser Thr Pro Gly Arg Ser Lys Ile
775 780 785

AAG TTT GAC ATG GAA AAA ATG CAC TCG GCT GTT GGG CAA GGT GTG CCA 2576
Lys Phe Asp Met Glu Lys Met His Ser Ala Val Gly Gln Gly Val Pro
790 795 800

CGT CAT CAC CGA GGT GAA ATC TGG AAA TTT CTA GCT GAG CAA TTC CAC 2624
Arg His His Arg Gly Glu Ile Trp Lys Phe Leu Ala Glu Gln Phe His
805 810 815

CTT AAA CAC CAG TTT CCC AGC AAA CAG CAG CCA AAG GAT GTG CCA TAC 2672
Leu Lys His Gln Phe Pro Ser Lys Gln Gln Pro Lys Asp Val Pro Tyr
820 825 830

AAA GAA CTC TTA AAG CAG CTG ACT TCC CAG CAG CAT GCG ATT CTT ATT 2720
Lys Glu Leu Leu Lys Gln Leu Thr Ser Gln Gln His Ala Ile Leu Ile
835 840 845 850

GAC CTT GGG CGA ACC TTT CCT ACA CAC CCA TAC TTC TCT GCC CAG CTT 2768
Asp Leu Gly Arg Thr Phe Pro Thr His Pro Tyr Phe Ser Ala Gln Leu
855 860 865

GGA GCA GGA CAG CTA TCG CTT TAC AAC ATT TTG AAG GCC TAC TCA CTT 2816
Gly Ala Gly Gln Leu Ser Leu Tyr Asn Ile Leu Lys Ala Tyr Ser Leu
870 875 880

```
CTA GAC CAG GAA GTG GGA TAT TGC CAA GGT CTC AGC TTT GTA GCA GGC      2864
Leu Asp Gln Glu Val Gly Tyr Cys Gln Gly Leu Ser Phe Val Ala Gly
        885                 890                 895

ATT TTG CTT CTT CAT ATG AGT GAG GAA GAG GCG TTT AAA ATG CTC AAG      2912
Ile Leu Leu Leu His Met Ser Glu Glu Glu Ala Phe Lys Met Leu Lys
    900                 905                 910

TTT CTG ATG TTT GAC ATG GGG CTG CGG AAA CAG TAT CGG CCA GAC ATG      2960
Phe Leu Met Phe Asp Met Gly Leu Arg Lys Gln Tyr Arg Pro Asp Met
915                 920                 925                 930

ATT ATT TTA CAG ATC CAG ATG TAC CAG CTC TCG AGG TTG CTT CAT GAT      3008
Ile Ile Leu Gln Ile Gln Met Tyr Gln Leu Ser Arg Leu Leu His Asp
                935                 940                 945

TAC CAC AGA GAC CTC TAC AAT CAC CTG GAG GAG CAC GAG ATC GGC CCC      3056
Tyr His Arg Asp Leu Tyr Asn His Leu Glu Glu His Glu Ile Gly Pro
            950                 955                 960

AGC CTC TAC GCT GCC CCC TGG TTC CTC ACC ATG TTT GCC TCA CAG TTC      3104
Ser Leu Tyr Ala Ala Pro Trp Phe Leu Thr Met Phe Ala Ser Gln Phe
        965                 970                 975

CCG CTG GGA TTC GTA GCC AGA GTC TTT GAT ATG ATT TTT CTT CAG GGA      3152
Pro Leu Gly Phe Val Ala Arg Val Phe Asp Met Ile Phe Leu Gln Gly
    980                 985                 990

ACA GAG GTC ATA TTT AAA GTG GCT TTA AGT CTG TTG GGA AGC CAT AAG      3200
Thr Glu Val Ile Phe Lys Val Ala Leu Ser Leu Leu Gly Ser His Lys
995                 1000                1005                1010

CCC TTG ATT CTG CAG CAT GAA AAC CTA GAA ACC ATA GTT GAC TTT ATA      3248
Pro Leu Ile Leu Gln His Glu Asn Leu Glu Thr Ile Val Asp Phe Ile
                1015                1020                1025

AAA AGC ACG CTA CCC AAC CTT GGC TTG GTA CAG ATG GAA AAG ACC ATC      3296
Lys Ser Thr Leu Pro Asn Leu Gly Leu Val Gln Met Glu Lys Thr Ile
            1030                1035                1040

AAT CAG GTA TTT GAA ATG GAC ATC GCT AAA CAG TTA CAA GCT TAT GAA      3344
Asn Gln Val Phe Glu Met Asp Ile Ala Lys Gln Leu Gln Ala Tyr Glu
        1045                1050                1055

GTT GAG TAC CAC GTC CTT CAA GAA GAA CTT ATC GAT TCC TCT CCT CTC      3392
Val Glu Tyr His Val Leu Gln Glu Glu Leu Ile Asp Ser Ser Pro Leu
    1060                1065                1070

AGT GAC AAC CAA AGA ATG GAT AAA TTA GAG AAA ACC AAC AGC AGC TTA      3440
Ser Asp Asn Gln Arg Met Asp Lys Leu Glu Lys Thr Asn Ser Ser Leu
1075                1080                1085                1090
```

```
CGC AAA CAG AAC CTT GAC CTC CTT GAA CAG TTG CAG GTG GCA AAT GGT      3488
Arg Lys Gln Asn Leu Asp Leu Leu Glu Gln Leu Gln Val Ala Asn Gly
                1095                1100                1105

AGG ATC CAA AGC CTT GAG GCC ACC ATT GAG AAG CTC CTG AGC AGT GAG      3536
Arg Ile Gln Ser Leu Glu Ala Thr Ile Glu Lys Leu Leu Ser Ser Glu
            1110                1115                1120

AGC AAG CTG AAG CAG GCC ATG CTT ACC TTA GAA CTG GAG CGG TCG GCC      3584
Ser Lys Leu Lys Gln Ala Met Leu Thr Leu Glu Leu Glu Arg Ser Ala
        1125                1130                1135

CTG CTG CAG ACG GTG GAG GAG CTG CGG CGG CGG AGC GCA GAG CCC AGC      3632
Leu Leu Gln Thr Val Glu Glu Leu Arg Arg Arg Ser Ala Glu Pro Ser
    1140                1145                1150

GAC CGG GAG CCT GAG TGC ACG CAG NCC CGA GCC CAC GGG CGA CTG ACA      3680
Asp Arg Glu Pro Glu Cys Thr Gln Xaa Arg Ala His Gly Arg Leu Thr
1155                1160                1165                1170

GCT CTG CAG GAG AGA TTG CAA CAC CAT CCC ACA CTG TCC AGG CCT TAA      3728
Ala Leu Gln Glu Arg Leu Gln His His Pro Thr Leu Ser Arg Pro  *
                1175                1180                1185

CTGAGAGGGA CAGAAGACGC TGGAAGGAGA GAAGGAAGCG GGAAGTGTGC TTCTCAGGGA   3788

GGAAACCGGC TTGCCAGCAA GTAGATTCTT ACGAACTCCA ACTTGCAATT CAGGGGGCAT   3848

GTCCCAGTGT TTTTTTTGTT GTTTTTAGAT ACTAAATCGT CCCTTCTCCA GTCCTGATTA   3908

CTGTACACAG TAGCTTTAGA TGGCGTGGAC GTGAATAAAT GCAACTTATG TTTTAAAAAA   3968

AAAAAAAAAA AAAAAA                                                   3984
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3989 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Homo sapiens

(ix) FEATURE:
- (A) NAME/KEY: homology with EST in :SEQID6
- (B) LOCATION: 1..458

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
ATAATAGGCA CTGAAGACAT GTTAATGGAA GGTGGATTTG TGATTCAGAA CCTCTAGACT      60

ACCTGGGCGA GTCTTTTAAA ATGTTTCTGC ATATGAAGTG TGTAAAATAG ATTGCTTGAT     120

CCAAAACAGA AAAACAGTGA TAACTGTTTT GCTGAGTTCC CAGACCCTTC CCAAG ATG      178
                                                              Met
                                                               1

GAA CCA ATA ACA TTC ACA GCA AGG AAA CAT CTG CTT CCT AAC GAG GTC        226
Glu Pro Ile Thr Phe Thr Ala Arg Lys His Leu Leu Pro Asn Glu Val
            5                  10                  15

TCG GTG GAT TTT GGC CTG CAG CTG GTG GGC TCC CTG CCT GTG CAT TCC        274
Ser Val Asp Phe Gly Leu Gln Leu Val Gly Ser Leu Pro Val His Ser
        20                  25                  30

CTG ACC ACC ATG CCC ATG CTG CCC TGG GTT GTG GCT GAG GTG CGA AGA        322
Leu Thr Thr Met Pro Met Leu Pro Trp Val Val Ala Glu Val Arg Arg
    35                  40                  45

CTC AGC AGG CAG TCC ACC AGA AAG GAA CCT GTA ACC AAG CAA GTC CGG        370
Leu Ser Arg Gln Ser Thr Arg Lys Glu Pro Val Thr Lys Gln Val Arg
50                  55                  60                  65

CTT TGC GTT TCA CCC TCT GGA CTG AGA TGT GAA CCT GAG CCA GGG AGA        418
Leu Cys Val Ser Pro Ser Gly Leu Arg Cys Glu Pro Glu Pro Gly Arg
                70                  75                  80

AGT CAA CAG TGG GAT CCC CTG ATC TAT TCC AGC ATC TTT GAG TGC AAG        466
Ser Gln Gln Trp Asp Pro Leu Ile Tyr Ser Ser Ile Phe Glu Cys Lys
            85                  90                  95

CCT CAG CGT GTT CAC AAA CTG ATT CAC AAC AGT CAT GAC CCA AGT TAC        514
Pro Gln Arg Val His Lys Leu Ile His Asn Ser His Asp Pro Ser Tyr
        100                 105                 110

TTT GCT TGT CTG ATT AAG GAA GAC GCT GTC CAC CGG CAG AGT ATC TGC        562
Phe Ala Cys Leu Ile Lys Glu Asp Ala Val His Arg Gln Ser Ile Cys
    115                 120                 125

TAT GTG TTC AAA GCC GAT GAT CAA ACA AAA GTG CCT GAG ATC ATC AGC        610
Tyr Val Phe Lys Ala Asp Asp Gln Thr Lys Val Pro Glu Ile Ile Ser
130                 135                 140                 145

TCC ATC CGT CAG GCG GGG AAG ATC GCC CGG CAG GAG GAG CTG CAC TGC        658
Ser Ile Arg Gln Ala Gly Lys Ile Ala Arg Gln Glu Glu Leu His Cys
                150                 155                 160
```

CCG TCC GAG TTC GAC GAC ACG TTT TCC AAG AAG TTC GAG GTG CTC TTC 706
Pro Ser Glu Phe Asp Asp Thr Phe Ser Lys Lys Phe Glu Val Leu Phe
165 170 175

TGC GGC CGC GTG ACG GTG GCG CAC AAG AAG GCT CCG CCG GCC CTG ATC 754
Cys Gly Arg Val Thr Val Ala His Lys Lys Ala Pro Pro Ala Leu Ile
180 185 190

GAC GAG TGC ATC GAG AAG TTC AAT CAC GTC AGC GGC AGC CGG GGG TCC 802
Asp Glu Cys Ile Glu Lys Phe Asn His Val Ser Gly Ser Arg Gly Ser
195 200 205

GAG AGC CCC CGC CCC AAC CCG CCC CAT GCC GCG CCC ACA GGG AGC CAG 850
Glu Ser Pro Arg Pro Asn Pro Pro His Ala Ala Pro Thr Gly Ser Gln
210 215 220 225

GAG CCT GTG CGC AGG CCC ATG CGC AAG TCC TTC TCC CAG CCC GGC CTG 898
Glu Pro Val Arg Arg Pro Met Arg Lys Ser Phe Ser Gln Pro Gly Leu
230 235 240

CGC TCG CTG GCC TTT AGG AAG GAG CTG CAG GAT GGG GGC CTC CGA AGC 946
Arg Ser Leu Ala Phe Arg Lys Glu Leu Gln Asp Gly Gly Leu Arg Ser
245 250 255

AGC GGC TTC TTC AGC TCC TTC GAG GAG AGC GAC ATT GAG AAC CAC CTC 994
Ser Gly Phe Phe Ser Ser Phe Glu Glu Ser Asp Ile Glu Asn His Leu
260 265 270

ATT AGC GGA CAC AAT ATT GTG CAG CCC ACA GAT ATC GAG GAA AAT CGA 1042
Ile Ser Gly His Asn Ile Val Gln Pro Thr Asp Ile Glu Glu Asn Arg
275 280 285

ACT ATG CTC TTC ACG ATT GGC CAG TCT GAA GTT TAC CTC ATC AGT CCT 1090
Thr Met Leu Phe Thr Ile Gly Gln Ser Glu Val Tyr Leu Ile Ser Pro
290 295 300 305

GAC ACC AAA AAA ATA GCA TTG GAG AAA AAT TTT AAG GAG ATA TCC TTT 1138
Asp Thr Lys Lys Ile Ala Leu Glu Lys Asn Phe Lys Glu Ile Ser Phe
310 315 320

TGC TCT CAG GGC ATC AGA CAC GTG GAC CAC TTT GGG TTT ATC TGT CGG 1186
Cys Ser Gln Gly Ile Arg His Val Asp His Phe Gly Phe Ile Cys Arg
325 330 335

GAG TCT TCC GGA GGT GGC GGC TTT CAT TTT GTC TGT TAC GTG TTT CAG 1234
Glu Ser Ser Gly Gly Gly Gly Phe His Phe Val Cys Tyr Val Phe Gln
340 345 350

TGC ACA AAT GAG GCT CTG GTT GAT GAA ATT ATG ATG ACC CTG AAA CAG 1282
Cys Thr Asn Glu Ala Leu Val Asp Glu Ile Met Met Thr Leu Lys Gln
355 360 365

GCC TTC ACG GTG GCC GCA GTG CAG CAG ACA GCT AAG GCG CCA GCC CAG 1330
Ala Phe Thr Val Ala Ala Val Gln Gln Thr Ala Lys Ala Pro Ala Gln
370 375 380 385

CTG TGT GAG GGC TGC CCC CTG CAA AGC CTG CAC AAG CTC TGT GAG AGG 1378
Leu Cys Glu Gly Cys Pro Leu Gln Ser Leu His Lys Leu Cys Glu Arg
390 395 400

ATA GAG GGA ATG AAT TCT TCC AAA ACA AAA CTA GAA CTG CAA AAG CAC 1426
Ile Glu Gly Met Asn Ser Ser Lys Thr Lys Leu Glu Leu Gln Lys His
405 410 415

CTG ACG ACA TTA ACC AAT CAG GAG CAG GCG ACT ATT TTT GAA GAG GTT 1474
Leu Thr Thr Leu Thr Asn Gln Glu Gln Ala Thr Ile Phe Glu Glu Val
420 425 430

CAG AAA TTG AGA CCG AGA AAT GAG CAG CGA GAG AAT GAA TTG ATT ATT 1522
Gln Lys Leu Arg Pro Arg Asn Glu Gln Arg Glu Asn Glu Leu Ile Ile
435 440 445

TCT TTT CTG AGA TGT TTA TAT GAA GAG AAA CAG AAA GAA CAC ATC CAT 1570
Ser Phe Leu Arg Cys Leu Tyr Glu Glu Lys Gln Lys Glu His Ile His
450 455 460 465

ATT GGG GAG ATG AAG CAG ACA TCG CAG ATG GCA GCA GAG AAT ATT GGA 1618
Ile Gly Glu Met Lys Gln Thr Ser Gln Met Ala Ala Glu Asn Ile Gly
470 475 480

AGT GAA TTA CCA CCC AGT GCC ACT CGA TTT AGG CTA GAT ATG CTG AAA 1666
Ser Glu Leu Pro Pro Ser Ala Thr Arg Phe Arg Leu Asp Met Leu Lys
485 490 495

AAC AAA GCA AAG AGA TCT TTA ACA GAG TCT TTA GAA AGT ATT TTG TCC 1714
Asn Lys Ala Lys Arg Ser Leu Thr Glu Ser Leu Glu Ser Ile Leu Ser
500 505 510

CGG GGT AAT AAA GCC AGA GGC CTG CAG GAA CAC TCC ATC AGT GTG GAT 1762
Arg Gly Asn Lys Ala Arg Gly Leu Gln Glu His Ser Ile Ser Val Asp
515 520 525

CTG GAT AGC TCC CTG TCT AGT ACA TTA AGT AAC ACC AGC AAA GAG CCA 1810
Leu Asp Ser Ser Leu Ser Ser Thr Leu Ser Asn Thr Ser Lys Glu Pro
530 535 540 545

TCT GTG TGT GAA AAG GAG GCC TTG CCC ATC TCT GAG AGC TCC TTT AAG 1858
Ser Val Cys Glu Lys Glu Ala Leu Pro Ile Ser Glu Ser Ser Phe Lys
550 555 560

CTC CTC GGC TCC TCG GAG GAC CTG TCC AGT GAC TCG GAG AGT CAT CTC 1906
Leu Leu Gly Ser Ser Glu Asp Leu Ser Ser Asp Ser Glu Ser His Leu
565 570 575

GGA ACA GAG GTC ATA TTT AAA GTG GCT TTA AGT CTG TTG GGA AGC CAT 3202
Gly Thr Glu Val Ile Phe Lys Val Ala Leu Ser Leu Leu Gly Ser His
995 1000 1005

AAG CCC TTG ATT CTG CAG CAT GAA AAC CTA GAA ACC ATA GTT GAC TTT 3250
Lys Pro Leu Ile Leu Gln His Glu Asn Leu Glu Thr Ile Val Asp Phe
1010 1015 1020 1025

ATA AAA AGC ACG CTA CCC AAC CTT GGC TTG GTA CAG ATG GAA AAG ACC 3298
Ile Lys Ser Thr Leu Pro Asn Leu Gly Leu Val Gln Met Glu Lys Thr
1030 1035 1040

ATC AAT CAG GTA TTT GAA ATG GAC ATC GCT AAA CAG TTA CAA GCT TAT 3346
Ile Asn Gln Val Phe Glu Met Asp Ile Ala Lys Gln Leu Gln Ala Tyr
1045 1050 1055

GAA GTT GAG TAC CAC GTC CTT CAA GAA GAA CTT ATC GAT TCC TCT CCT 3394
Glu Val Glu Tyr His Val Leu Gln Glu Glu Leu Ile Asp Ser Ser Pro
1060 1065 1070

CTC AGT GAC AAC CAA AGA ATG GAT AAA TTA GAG AAA ACC AAC AGC AGC 3442
Leu Ser Asp Asn Gln Arg Met Asp Lys Leu Glu Lys Thr Asn Ser Ser
1075 1080 1085

TTA CGC AAA CAG AAC CTT GAC CTC CTT GAA CAG TTG CAG GTG GCA AAT 3490
Leu Arg Lys Gln Asn Leu Asp Leu Leu Glu Gln Leu Gln Val Ala Asn
1090 1095 1100 1105

GGT AGG ATC CAA AGC CTT GAG GCC ACC ATT GAG AAG CTC CTG AGC AGT 3538
Gly Arg Ile Gln Ser Leu Glu Ala Thr Ile Glu Lys Leu Leu Ser Ser
1110 1115 1120

GAG AGC AAG CTG AAG CAG GCC ATG CTT ACC TTA GAA CTG GAG CGG TCG 3586
Glu Ser Lys Leu Lys Gln Ala Met Leu Thr Leu Glu Leu Glu Arg Ser
1125 1130 1135

GCC CTG CTG CAG ACG GTG GAG GAG CTG CGG CGG CGG AGC GCA GAG CCC 3634
Ala Leu Leu Gln Thr Val Glu Glu Leu Arg Arg Arg Ser Ala Glu Pro
1140 1145 1150

AGC GAC CGG GAG CCT GAG TGC ACG CAG NCC CGA GCC CAC GGG CGA CTG 3682
Ser Asp Arg Glu Pro Glu Cys Thr Gln Xaa Arg Ala His Gly Arg Leu
1155 1160 1165

ACA GCT CTG CAG GAG AGA TTG CAA CAC CAT CCC ACA CTG TCC AGG CCT 3730
Thr Ala Leu Gln Glu Arg Leu Gln His His Pro Thr Leu Ser Arg Pro
1170 1175 1180 1185

TAA CTGAGAGGGA CAGAAGACGC TGGAAGGAGA GAAGGAAGCG GGAAGTGTGC 3783
*

TTCTCAGGGA GGAAACCGGC TTGCCAGCAA GTAGATTCTT ACGAACTCCA ACTTGCAATT 3843

CAGGGGGCAT GTCCCAGTGT TTTTTTTGTT GTTTTTAGAT ACTAAATCGT CCCTTCTCCA 3903

GTCCTGATTA CTGTACACAG TAGCTTTAGA TGGCGTGGAC GTGAATAAAT GCAACTTATG 3963

TTTTAAAAAA AAAAAAAAAA AAAAAA 3989

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1185 amino acids
(B) TYPE: AMINO ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Met Glu Pro Ile Thr Phe Thr Ala Arg Lys His Leu Leu Pro Asn Glu
1 5 10 15

Val Ser Val Asp Phe Gly Leu Gln Leu Val Gly Ser Leu Pro Val His
20 25 30

Ser Leu Thr Thr Met Pro Met Leu Pro Trp Val Val Ala Glu Val Arg
35 40 45

Arg Leu Ser Arg Gln Ser Thr Arg Lys Glu Pro Val Thr Lys Gln Val
50 55 60

Arg Leu Cys Val Ser Pro Ser Gly Leu Arg Cys Glu Pro Glu Pro Gly
65 70 75 80

Arg Ser Gln Gln Trp Asp Pro Leu Ile Tyr Ser Ser Ile Phe Glu Cys
85 90 95

Lys Pro Gln Arg Val His Lys Leu Ile His Asn Ser His Asp Pro Ser
100 105 110

Tyr Phe Ala Cys Leu Ile Lys Glu Asp Ala Val His Arg Gln Ser Ile
115 120 125

Cys Tyr Val Phe Lys Ala Asp Asp Gln Thr Lys Val Pro Glu Ile Ile
130 135 140

Ser Ser Ile Arg Gln Ala Gly Lys Ile Ala Arg Gln Glu Glu Leu His
145 150 155 160

Cys Pro Ser Glu Phe Asp Asp Thr Phe Ser Lys Lys Phe Glu Val Leu
165 170 175

Phe Cys Gly Arg Val Thr Val Ala His Lys Lys Ala Pro Pro Ala Leu
180 185 190

Ile Asp Glu Cys Ile Glu Lys Phe Asn His Val Ser Gly Ser Arg Gly
195 200 205

Ser Glu Ser Pro Arg Pro Asn Pro Pro His Ala Ala Pro Thr Gly Ser
210 215 220

Gln Glu Pro Val Arg Arg Pro Met Arg Lys Ser Phe Ser Gln Pro Gly
225 230 235 240

Leu Arg Ser Leu Ala Phe Arg Lys Glu Leu Gln Asp Gly Gly Leu Arg
245 250 255

Ser Ser Gly Phe Phe Ser Ser Phe Glu Glu Ser Asp Ile Glu Asn His
260 265 270

Leu Ile Ser Gly His Asn Ile Val Gln Pro Thr Asp Ile Glu Glu Asn
275 280 285

Arg Thr Met Leu Phe Thr Ile Gly Gln Ser Glu Val Tyr Leu Ile Ser
290 295 300

Pro Asp Thr Lys Lys Ile Ala Leu Glu Lys Asn Phe Lys Glu Ile Ser
305 310 315 320

Phe Cys Ser Gln Gly Ile Arg His Val Asp His Phe Gly Phe Ile Cys
325 330 335

Arg Glu Ser Ser Gly Gly Gly Gly Phe His Phe Val Cys Tyr Val Phe
340 345 350

Gln Cys Thr Asn Glu Ala Leu Val Asp Glu Ile Met Met Thr Leu Lys
355 360 365

Gln Ala Phe Thr Val Ala Ala Val Gln Gln Thr Ala Lys Ala Pro Ala
370 375 380

Gln Leu Cys Glu Gly Cys Pro Leu Gln Ser Leu His Lys Leu Cys Glu
385 390 395 400

Arg Ile Glu Gly Met Asn Ser Ser Lys Thr Lys Leu Glu Leu Gln Lys
405 410 415

His Leu Thr Thr Leu Thr Asn Gln Glu Gln Ala Thr Ile Phe Glu Glu
420 425 430

Val Gln Lys Leu Arg Pro Arg Asn Glu Gln Arg Glu Asn Glu Leu Ile
435 440 445

Ile Ser Phe Leu Arg Cys Leu Tyr Glu Glu Lys Gln Lys Glu His Ile
450 455 460

His Ile Gly Glu Met Lys Gln Thr Ser Gln Met Ala Ala Glu Asn Ile
465 470 475 480

Gly Ser Glu Leu Pro Pro Ser Ala Thr Arg Phe Arg Leu Asp Met Leu
485 490 495

Lys Asn Lys Ala Lys Arg Ser Leu Thr Glu Ser Leu Glu Ser Ile Leu
500 505 510

Ser Arg Gly Asn Lys Ala Arg Gly Leu Gln Glu His Ser Ile Ser Val
515 520 525

Asp Leu Asp Ser Ser Leu Ser Ser Thr Leu Ser Asn Thr Ser Lys Glu
530 535 540

Pro Ser Val Cys Glu Lys Glu Ala Leu Pro Ile Ser Glu Ser Ser Phe
545 550 555 560

Lys Leu Leu Gly Ser Ser Glu Asp Leu Ser Ser Asp Ser Glu Ser His
565 570 575

Leu Pro Glu Glu Pro Ala Pro Leu Ser Pro Gln Gln Ala Phe Arg Arg
580 585 590

Arg Ala Asn Thr Leu Ser His Phe Pro Ile Glu Cys Gln Glu Pro Pro
595 600 605

Gln Pro Ala Arg Gly Ser Pro Gly Val Ser Gln Arg Lys Leu Met Arg
610 615 620

Tyr His Ser Val Ser Thr Glu Thr Pro His Glu Arg Lys Asp Phe Glu
625 630 635 640

Ser Lys Ala Asn His Leu Gly Asp Ser Gly Gly Thr Pro Val Lys Thr
645 650 655

Arg Arg His Ser Trp Arg Gln Gln Ile Phe Leu Arg Val Ala Thr Pro
660 665 670

Gln Lys Ala Cys Asp Ser Ser Ser Arg Tyr Glu Asp Tyr Ser Glu Leu
675 680 685

Gly Glu Leu Pro Pro Arg Ser Pro Leu Glu Pro Val Cys Glu Asp Gly
690 695 700

Pro Phe Gly Pro Pro Pro Glu Glu Lys Lys Arg Thr Ser Arg Glu Leu
705 710 715 720

Arg Glu Leu Trp Gln Lys Ala Ile Leu Gln Gln Ile Leu Leu Leu Arg
725 730 735

Met Glu Lys Glu Asn Gln Lys Leu Gln Ala Ser Glu Asn Asp Leu Leu
740 745 750

Asn Lys Arg Leu Lys Leu Asp Tyr Glu Glu Ile Thr Pro Cys Leu Lys
755 760 765

Glu Val Thr Thr Val Trp Glu Lys Met Leu Ser Thr Pro Gly Arg Ser
770 775 780

Lys Ile Lys Phe Asp Met Glu Lys Met His Ser Ala Val Gly Gln Gly
785 790 795 800

Val Pro Arg His His Arg Gly Glu Ile Trp Lys Phe Leu Ala Glu Gln
805 810 815

Phe His Leu Lys His Gln Phe Pro Ser Lys Gln Gln Pro Lys Asp Val
820 825 830

Pro Tyr Lys Glu Leu Leu Lys Gln Leu Thr Ser Gln Gln His Ala Ile
835 840 845

Leu Ile Asp Leu Gly Arg Thr Phe Pro Thr His Pro Tyr Phe Ser Ala
850 855 860

Gln Leu Gly Ala Gly Gln Leu Ser Leu Tyr Asn Ile Leu Lys Ala Tyr
865 870 875 880

Ser Leu Leu Asp Gln Glu Val Gly Tyr Cys Gln Gly Leu Ser Phe Val
885 890 895

Ala Gly Ile Leu Leu Leu His Met Ser Glu Glu Glu Ala Phe Lys Met
900 905 910

Leu Lys Phe Leu Met Phe Asp Met Gly Leu Arg Lys Gln Tyr Arg Pro
915 920 925

Asp Met Ile Ile Leu Gln Ile Gln Met Tyr Gln Leu Ser Arg Leu Leu
930 935 940

His Asp Tyr His Arg Asp Leu Tyr Asn His Leu Glu Glu His Glu Ile
945 950 955 960

Gly Pro Ser Leu Tyr Ala Ala Pro Trp Phe Leu Thr Met Phe Ala Ser
965 970 975

Gln Phe Pro Leu Gly Phe Val Ala Arg Val Phe Asp Met Ile Phe Leu
980 985 990

Gln Gly Thr Glu Val Ile Phe Lys Val Ala Leu Ser Leu Leu Gly Ser
995 1000 1005

His Lys Pro Leu Ile Leu Gln His Glu Asn Leu Glu Thr Ile Val Asp
1010 1015 1020

Phe Ile Lys Ser Thr Leu Pro Asn Leu Gly Leu Val Gln Met Glu Lys
1025 1030 1035 1040

Thr Ile Asn Gln Val Phe Glu Met Asp Ile Ala Lys Gln Leu Gln Ala
1045 1050 1055

Tyr Glu Val Glu Tyr His Val Leu Gln Glu Glu Leu Ile Asp Ser Ser
1060 1065 1070

Pro Leu Ser Asp Asn Gln Arg Met Asp Lys Leu Glu Lys Thr Asn Ser
1075 1080 1085

Ser Leu Arg Lys Gln Asn Leu Asp Leu Leu Glu Gln Leu Gln Val Ala
1090 1095 1100

Asn Gly Arg Ile Gln Ser Leu Glu Ala Thr Ile Glu Lys Leu Leu Ser
1105 1110 1115 1120

Ser Glu Ser Lys Leu Lys Gln Ala Met Leu Thr Leu Glu Leu Glu Arg
1125 1130 1135

Ser Ala Leu Leu Gln Thr Val Glu Glu Leu Arg Arg Arg Ser Ala Glu
1140 1145 1150

Pro Ser Asp Arg Glu Pro Glu Cys Thr Gln Ser Arg Ala His Gly Arg
1155 1160 1165

Leu Thr Ala Leu Gln Glu Arg Leu Gln His His Pro Thr Leu Ser Arg
1170 1175 1180

Pro
1185

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 458 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: 5' EST

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

ATAATAGGCA CTGAAGACAT GTTAATGGAA GGTGGATTTG TGATTCAGAA CCTCTAGACT 60

ACCTGGGCGA GTCTTTTAAA ATGTTTCTGC ATATGAAGTG TGTAAAATAG ATTGCTTGAT 120

CCAAAACAGA AAAACAGTGA TAACTGTTTT GCTGAGTTCC CAGACCCTTC CCAAGATGGA 180

ACCAATAACA TTCACAGCAA GGAAACATCT GCTTCCTAAC GAGGTCTCGG TGGATTTTGG 240

CCTGCAGCTG GTGGGCTCCC TGCCTGTGCA TTCCCTGACC ACCATGCCCA TGCTGCCCTG 300

GGTTGTGGCT GAGGTGCGAA GACTCAGCAG GCAGTCCACC AGAAAGGAAC CTGTAACCAN 360

GCAANTCCGG CTTTGCGTTT CACCCTCTGG ACTGAGATGT GAACCTGAGC CAGGGAGAAG 420

TCAACAGTGG GATCCCCTGA TCTATTCCAG CATCTTTG 458

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: polymorphic fragment 99-430-352
(B) LOCATION: 1..47

(ix) FEATURE:
(A) NAME/KEY: polymorphic base
(B) LOCATION: 24
(D) OTHER INFORMATION: base c

(ix) FEATURE:
(A) NAME/KEY: Potential microsequencing oligo 99-430-352
(B) LOCATION: 1..23

(ix) FEATURE:
(A) NAME/KEY: Potential microsequencing oligo 99-430-352
(B) LOCATION: complement 25..47

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ACTTGCAAAC TAACACGTTA GCCCGCCACA GTTATCCTGC TGGTAGA 47

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: polymorphic fragment 99-430-352
(B) LOCATION: 1..47
(D) OTHER INFORMATION: variant version of SEQ ID7

(ix) FEATURE:
(A) NAME/KEY: polymorphic base
(B) LOCATION: 24
(D) OTHER INFORMATION: base t; c in SEQ ID7

(ix) FEATURE:
(A) NAME/KEY: Potential microsequencing oligo 99-430-352
(B) LOCATION: 1..23

(ix) FEATURE:
(A) NAME/KEY: Potential microsequencing oligo 99-430-352
(B) LOCATION: complement 25..47

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

ACTTGCAAAC TAACACGTTA GCCTGCCACA GTTATCCTGC TGGTAGA 47

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: upstream amplification primer for SEQ ID7, SEQ ID8
(B) LOCATION: 1..18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CACCCATCGC ATATTTCC 18

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: downstream amplification primer for SEQ ID7, SEQ ID8
(B) LOCATION: 1..18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AGAAATGCTG ACTGTCTG 18

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: microsequencing oligo 99-430-352.mis1
(B) LOCATION: 1..19

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GCAAACTAAC ACGTTAGCC 19

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: potential microsequencing oligo 99-430-352.mis2
(B) LOCATION: 1..23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TCTACCAGCA GGATAACTGT GGC 23

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: primer for cDNA isolation
(B) LOCATION: 1..19

(ix) FEATURE:
(A) NAME/KEY: positions in SEQID3
(B) LOCATION: 271..289

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TGACCACCAT GCCCATGCT 19

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: primer for cDNA isolation
(B) LOCATION: 1..21

(ix) FEATURE:
(A) NAME/KEY: positions in SEQID3
(B) LOCATION: complement 3950..3930

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GCATTTATTC ACGTCCACGC C 21